

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:47:23 ; Search time 4508.21 Seconds
(without alignments)
17679.337 Million cell updates/sec

Title: US-09-989-981A-7
Perfect score: 2669
Sequence: 1 gtgtccctgctccaggaaac.....caattaaaaatgtattgagc 2669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*

28: gb_gssl:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query						
No.	Score	Match	Length	DB	ID	Description		
	1	1511.6	56.6	3623	11	AK004871	AK004871 Mus muscu	
	2	1286.2	48.2	2417	11	AK050938	AK050938 Mus muscu	
	3	681.8	25.5	691	13	BX481838	BX481838 DKFZp686M	
	4	460.6	17.3	849	12	BI330745	BI330745 602982409	
	5	370.4	13.9	549	10	BF660076	BF660076 maa27c08.	
	6	361.4	13.5	583	13	BY705076	BY705076 BY705076	
c	7	355.8	13.3	457	14	T91380	T91380 yd53b02.s1	
	8	332.4	12.5	334	13	BX482362	BX482362 DKFZp686F	
	9	331.2	12.4	510	10	BB610072	BB610072 BB610072	
	10	323	12.1	511	9	AI157365	AI157365 ui45h01.y	
	11	318.4	11.9	564	14	T84531	T84531 yd53b02.rl	
	12	309.8	11.6	500	9	AI151811	AI151811 ui46cl0.y	
	13	276.8	10.4	463	9	AA537862	AA537862 vj35a03.r	
c	14	276	10.3	781	14	CB502603	CB502603 ssalmge50	
	15	263.2	9.9	640	14	CD739823	CD739823 4028769 1	
c	16	236	8.8	613	14	CF367733	CF367733 852301 MA	
	17	234.6	8.8	398	9	AI597406	AI597406 vj35a03.y	
	18	226.6	8.5	586	11	AK008188	AK008188 Mus muscu	
	19	226.4	8.5	581	13	BY708144	BY708144 BY708144	
c	20	202.8	7.6	435	13	BX099922	BX099922 BX099922	
	21	198.4	7.4	435	9	AI574075	AI574075 uj67h11.y	
c	22	198	7.4	762	29	CC659228	CC659228 OGUFF57TV	
	23	196.4	7.4	821	28	BZ650554	BZ650554 OGCBA89TC	
	24	194	7.3	916	29	CG323718	CG323718 OG0DQ45TH	
c	25	193.2	7.2	709	29	CC695831	CC695831 OGULO23TV	
	26	187.2	7.0	776	29	CG327545	CG327545 OGWFJ96TV	
c	27	183.8	6.9	891	29	CG368338	CG368338 OG3BP65TV	
	28	183.2	6.9	826	29	CG214497	CG214497 OG1BM08TV	
	29	180.4	6.8	578	14	CF366327	CF366327 840972 MA	
	30	171	6.4	849	29	CG270361	CG270361 OGWFS70TH	
	31	171	6.4	857	29	CG271003	CG271003 OG0EJ71TV	
c	32	166.8	6.2	861	29	CG262933	CG262933 OG1DH53TV	
	33	162.4	6.1	345	14	CD730599	CD730599 4038931 1	
	34	155.6	5.8	839	29	CG262656	CG262656 OG1AN46TH	
c	35	151	5.7	912	29	CC604602	CC604602 OGUFQ75TH	
	36	150	5.6	909	29	CG268466	CG268466 OG2BT15TH	
	37	149.2	5.6	523	9	AU195806	AU195806 AU195806	
	38	145.4	5.4	447	12	BI145065	BI145065 602909138	
	39	144.2	5.4	936	10	BF162656	BF162656 601769307	
	40	143.2	5.4	694	13	CA140253	CA140253 SCEZRT202	
	41	141.6	5.3	823	14	CB649273	CB649273 OSJNEb13B	
	42	141	5.3	833	10	BF620684	BF620684 HVSMEc002	
	43	138.8	5.2	566	9	AU192726	AU192726 AU192726	
	44	138	5.2	560	9	AU192604	AU192604 AU192604	
	45	137.8	5.2	754	14	CB627408	CB627408 OSIIEb02F	

ALIGNMENTS

RESULT 1

AK004871

LOCUS AK004871 3623 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300003C16 product:ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full insert sequence.

ACCESSION AK004871

VERSION AK004871.1 GI:12836380

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 3623)

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cdDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

source Location/Qualifiers

1. .3623

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:1300003C16"

/db_xref="MGI:1896857"

/db_xref="taxon:10090"

/clone="1300003C16"

/sex="male"

/tissue_type="liver"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

CDS

69. .2090

/note="unnamed protein product; ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus] (SWISSPROT|Q9DBM0, evidence: FASTY, 92%ID, 96.7%length, match=1796)

putative"

/codon_start=1

/protein_id="BAB23630.1"

/db_xref="GI:12836381"

/translation="MAEKTKEETQLWNGTVLQDASQGLQDSLFSSSDNSLYFTYSQ
SNTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSQML
AIIGSSGCGRASLLDVTGRGHGGMKSGQIWINQGPSTPQLVRKCVAVHRQHDQLLP
NLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGE
RRRVSIGVQLLWNPGLILDEPTSGLDSEFAHNLTTLRLAKGNRLVLI SLHQPRSD
IFRLFDLVLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKE
REVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTLTQDTCGTAVEL
PGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAA
LLFMIGALIPFNVILDVSKCHSERSMLYIELEDGLYTAGPYFFAKILGELPEHCAYV
IIYAMPIYWLTNLRPVPELFLHFLLVVLVVFCCRTMALAASAMLPTEHMSFFCNAL
YNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSI
LGDTMISAMDNLNHPYAIYLVIGISYGFLEFLYYLSLKLKQKSIQDW"

polyA_signal 3605. .3610
/note="putative"
polyA_site 3623
/note="putative"

ORIGIN

Query Match 56.6%; Score 1511.6; DB 11; Length 3623;
Best Local Similarity 77.0%; Pred. No. 0;
Matches 1965; Conservative 0; Mismatches 534; Indels 53; Gaps 8;

Qy	99	CATGGCCGGGAAGGCGGCAGAGGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATAC	158
Db	68	CATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGC	127
Qy	159	CTC---GGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCAC	215
Db	128	TTCGCAGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCAC	187
Qy	216	CTACAGTGGCCAGCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGC	275
Db	188	CTACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGC	247
Qy	276	CTCTCAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAG	335
Db	248	CTCTCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAG	307
Qy	336	CTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCA	395
Db	308	CAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACA	367
Qy	396	GATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCAC	455
Db	368	GATGCTGGCCATCATAGGGAGCTCAGGTCGCGGAGAGCCTCACTACTCGACGTGATCAC	427
Qy	456	TGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAG	515
Db	428	AGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAG	487
Qy	516	CTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCC	575
Db	488	TACGCCTCAGCTGGTGAGGAAGTGCCTTGCATGTGCGGCAGCATGACCAACTGCTGCC	547
Qy	576	CAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCAGATGCGGCTGCCAGAACCTT	635
Db	548	CAACCTGACCGTCAGAGAGACCTGGCTTTCATTGCCAGATGCGCCTGCCAGGACCTT	607

Qy 636 CTCCCAGGCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCA 695
 |||||
 Db 608 CTCCCAGGCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCA 667

Qy 696 GTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTCTGGGGGGTGAGCGCAG 755
 |||||
 Db 668 GTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCG 727

Qy 756 GAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACC 815
 |||||
 Db 728 ACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACC 787

Qy 816 CACCTCTGGGCTCGACAGCTTCACAGCCCAACCTGGTGAAGACCTTGTCCAGGCTGGC 875
 |||||
 Db 788 CACTTCTGGCCTCGACAGCTTCACAGCCCAACCTGTTGACAACCTTGTCCCGCCTGGC 847

Qy 876 CAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCT 935
 |||||
 Db 848 CAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCT 907

Qy 936 GTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCA 995
 |||||
 Db 908 ATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCA 967

Qy 996 CATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGA 1055
 |||||
 Db 968 AATGGTGCAGTACTTCACATCCATTGGCCACCCTGTCTCGCTATAGCAACCCTGCGGA 1027

Qy 1056 CTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAG 1115
 |||||
 Db 1028 CTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGT 1087

Qy 1116 GGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGGGTGACTTAGATGACTT 1175
 |||||
 Db 1088 GGAGAAGGCACAGTCTCTTGACGCCCTGTCTCTAGAAAAAGTACAAGGCTTTGATGACTT 1147

Qy 1176 TCTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGAAAGCAGCGTGAC 1235
 |||||
 Db 1148 TCTGTGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCT 1207

Qy 1236 CCCACTAGACACCAACTGCCTCCCGAGTCTACGAAGATGCCTGGGGCGGTGACGAGTT 1295
 |||||
 Db 1208 CACACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTT 1264

Qy 1296 TACGACGCTGATCCGTCGTGAGATTTCCAACGACTTCCGAGACCTGCCCCACCTCCTCAT 1355
 |||||
 Db 1265 TTCCACCCTGATCCGTCGTGAGATTTCCAATGACTTCCGGGACCTGCCCCAGCTGCTCAT 1324

Qy 1356 CCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGG 1415
 |||||
 Db 1325 TCATGGGTGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGG 1384

Qy 1416 GAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCAT 1475
 |||||
 Db 1385 GGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCAT 1444

Qy	1476	CCCTTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTA	1535
Db	1445		
		TCCTTTCAATGTCATCCTGGATGTCGTCCTCCAAATGTCACCTCGGAGAGGTCAATGCTGTA	1504
Qy	1536	CTATGAAGTGGGAGGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGG	1595
Db	1505		
		CTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCCCTTATTTCTTTGCCAAGATCCTAGG	1564
Qy	1596	GGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGC	1655
Db	1565		
		AGAATTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGAC	1624
Qy	1656	CAACCTGAGGCCAGGCCTCCAGCCCTTCTTGCTGCACTTCTTGCTGGTGTGGCTGGTGGT	1715
Db	1625		
		AAACCTGCGGCCCGTGCCTGAGCTCTTCCTTCTACACTTCTTGCTCGTGTGGTTGGTGGT	1684
Qy	1716	CTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGC	1775
Db	1685		
		CTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTC	1744
Qy	1776	CTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAA	1835
Db	1745		
		CTCCTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAA	1804
Qy	1836	CTTGAGCAGCCTGTGGACAGTGGCCGCTGGATTTCCAAAGTGTCTTCTGCGGTGGTG	1895
Db	1805		
		CTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTGCTTCTCCGGTGGTG	1864
Qy	1896	TTTTGAAGGGCTGATGAAGATTCAAGTTCAGTCAGAGAAGAACTTATAAAATGCCTCTCGGGAA	1955
Db	1865		
		CTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAA	1924
Qy	1956	CCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCC	2015
Db	1925		
		CTTCACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCC	1984
Qy	2016	TCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCTGTACTA	2075
Db	1985		
		ACTCTATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCTGTTCCTGTACTA	2044
Qy	2076	CGTGTCTTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGATTACAGCCAGACGT	2135
Db	2045		
		TCTATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGATACTCAGCCTTGCT	2104
Qy	2136	CTGCCCCGCTGGTGGGGGACCTGAGCAGACCTTCAACTGCACTCCCTCCTCAGGAGCCCC	2195
Db	2105		
		CTCACTGGCGG-----GACCCCTTTTCCCGGGGCTGGCCACCCAGGAGGAGCC	2152
Qy	2196	TTCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACATCCGGCCCAGGGTG	2255
Db	2153		
		GGACTGGGGACAAGGCTCACACAGATCTCTCAG-----GCAGCAGCCACCTCTTAGTG	2205
Qy	2256	CTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAGTCGAAAGGGATTT	2315
Db	2206		
		CTGCAGTGGCACAGGTGAGCCACAGGATGGCAGTAGAATAAAGACAGTTGAGAGGTGTTT	2265
Qy	2316	CTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTCGGTGGCACCTACA	2375

```

      ||||| | || | || |||| | |||| || |||
Db      2266 CTGCTCCCAGGCCAGGCTTGTGATGGGAGAGAGAGAA-----ACCAGGT 2310
Qy      2376 ACGTTGCTAATTTATTTCTTTTGATATGCATTTATATAGGCAACTCGATATAGGATGGG 2435
      ||||| || ||| ||| ||| ||| ||| ||| ||| |||
Db      2311 ACGTTGCTCATGCATTT-----TATATCTTTAAATAAACAACCCAGTATGGAATGGG 2362
Qy      2436 AGCAAAC TAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGA 2495
      | | | | | ||||| ||||| || |||| ||| |||| | ||| |
Db      2363 AACCAATTATATATGAATTGAGTAGCTAGGCTATGCAGAAATTTCTGGAATCCTGAGAGG 2422
Qy      2496 ACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCAGGGGCCCCACACTCCGTGGTGA 2555
      | | | | ||| ||| |||| | | |||| ||| |
Db      2423 ATAGTGGTTTATAGCAAAGTGTTAACTTTCTCTTCTACCATTCTCACAC----TGTAA 2478
Qy      2556 GCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATG-CCATCCCTTCTTTT 2614
      |||| | |||| | | ||||| | || |||| ||| || ||| |
Db      2479 GCCACTCCCAATACAAAGGGCGACCTAAACAAACTAGCAAAATGTTTTTCGCTTATCTC 2538
Qy      2615 TGTGTGGGGTCATGGGCTCCAAAAGCCAACGT 2646
      || |||| ||||| ||||| |||| ||
Db      2539 TCGGTGGATTTCATGGACTCCAACCCCCAAAGT 2570

```

RESULT 2

AK050938

LOCUS AK050938 2417 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030040P06 product:ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full insert sequence.

ACCESSION AK050938

VERSION AK050938.1 GI:26094211

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 2417)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/.

FEATURES Location/Qualifiers

source 1. .2417
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D030040P06"
/db_xref="MGI:2418860"
/db_xref="taxon:10090"
/clone="D030040P06"

```

/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9 days embryo"
misc_feature 1. .2417
/note="ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8
(STEROLIN-2) homolog [Mus musculus] (SWISSPROT|Q9DBM0,
evidence: FASTY, 92%ID, 96.7%length, match=1796)"

```

ORIGIN

```

Query Match          48.2%;  Score 1286.2;  DB 11;  Length 2417;
Best Local Similarity 76.1%;  Pred. No. 1e-286;
Matches 1696;  Conservative 0;  Mismatches 483;  Indels 50;  Gaps 7;

```

```

Qy      419 CAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGA 478
      |||| || ||||| ||||| || || || ||||| |||| |||| ||||| |||||
Db      184 CAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGA 243

Qy      479 TCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAGT 538
      | || ||||| || || ||||| ||||| ||||| ||||| ||||| |||||
Db      244 TGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGT 303

Qy      539 GTGTGGCCACGTGCGCCAGCACCAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCT 598
      | || || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      304 GCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCAACCTGACCGTCAGAGAGACCC 363

Qy      599 TGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCGTGACAAAA 658
      |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      364 TGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCAGCGTGACAAAC 423

Qy      659 GGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCA 718
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      424 GGGTGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCA 483

Qy      719 ACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGC 778
      ||| ||| || || || |||| ||||| ||||| ||||| ||||| |||||
Db      484 ACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGC 543

Qy      779 TCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCA 838
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      544 TCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCA 603

Qy      839 CAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCA 898
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      604 CAGCCCACAATCTGGTGACAACCTTGTCCGCTGGCCAAGGGCAACAGGCTGGTGCTCA 663

Qy      899 TCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGA 958
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      664 TCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGA 723

Qy      959 CGTCTGGCACCCCATCTACTTAGGGGCGGCCCAGCACATGGTCCAGTATTTACAGCCA 1018
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      724 CATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCA 783

Qy      1019 TCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCA 1078
      | ||| |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      784 TTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCA 843

```

Qy 1079 TTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAG 1138
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 844 TCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCCTGGGAGAAGGCACAGTCTCTTGCGAG 903

Qy 1139 CCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGG 1198
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 904 CCCTGTTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTGTGGAAAGCTGAGGCAAAGG 963

Qy 1199 ATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCACTAGACACCAACTGCCTCC 1258
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 964 AACTCAACACAAGCACCCACACAGTCAGCCTGACCCCTCACACAGGACACTGACTG---TG 1020

Qy 1259 CGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGA 1318
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1021 GGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTCCACCCTGATCCGTCGTCAGA 1080

Qy 1319 TTTCCAACGACTTCCGAGACCTGCCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGA 1378
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1081 TTTCCAATGACTTCCGGGACCTGCCCCACGCTGCTCATTCATGGGTGCGAAGCCTGCCTGA 1140

Qy 1379 TGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGG 1438
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1141 TGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAGCAGCTCTCCTTCATGG 1200

Qy 1439 ATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATG 1498
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1201 ACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTCAATGTCATCTGGATG 1260

Qy 1499 TCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGT 1558
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1261 TCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGT 1320

Qy 1559 ACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCT 1618
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1321 ACATGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCT 1380

Qy 1619 ACATCATCATCTACGGGATGCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCCTCCAGC 1678
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1381 ACGTCATCATCTACGCGATGCCATCTACTGGCTGACAAACCTGCGGCCCGTGCCTGAGC 1440

Qy 1679 CCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCC 1738
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1441 TCTTCCTTCTACACTTCCTGCTCGTGTGGTGGTGGTCTTCTGCTGCAGGACCATGGCCC 1500

Qy 1739 TGGCCGCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAAGCAATGCCCTCT 1798
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1501 TGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCTCCTTCTTCTGCAATGCCCTCT 1560

Qy 1799 ACAACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGC 1858
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1561 ACAACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGAGCAACCTGTGGATAGTGC 1620

Qy 1859 CCGCGTGGATTTCCAAAGTGTCTTCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTC 1918
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1621 CTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGTCTTCGGGGCTGATGCAGATTC 1680

Qy 1919 AGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAG 1978

Db	1681	AATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAG	1740
Qy	1979	ATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCG	2038
Db	1741	ACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTG	1800
Qy	2039	TCATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCCTTAAGGTTTCATCAAAC	2098
Db	1801	TCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTATCCTTGAAGCTCATCAAAC	1860
Qy	2099	AGAAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCCTGGTGGGGGACCTGA	2158
Db	1861	AGAAGTCAATTCAAGACTGGTGATACTCAGCCTTGCTCTCACTGGCGG-----	1908
Qy	2159	GCAGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAAT	2218
Db	1909	GACCCTTTTCCCGGGGCTGGCCACCCAGGAGGAGCCGGACTGGGGACAAGGCTCACACA	1968
Qy	2219	GACCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCAC	2278
Db	1969	GATCTCTCAG-----GCAGCAGCCACCTCTTAGTGCTGCAGTGGCACAGGTGAGCCAC	2021
Qy	2279	AGGATGGCAGTAGAATAAAGACAGTCGAAAGGGATTCTGCTCACTGGCAGGAGACTGCG	2338
Db	2022	AGGATGGCAGTAGAATAAAGACAGTTGAGAGGTGTTTCTGCTCCCAGGCCCAGGCTTGTTG	2081
Qy	2339	ATGACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTT	2398
Db	2082	ATGGGAGAGAGAGAA-----ACCAGGTACGTTGCTCATGCATTT-----	2120
Qy	2399	GATATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAACTAGGAATGAATTGGGT	2458
Db	2121	--TATATCTTTAAATAAACAACCCAGTATGGAATGGGAACCAATTATATATGAATTGAGT	2178
Qy	2459	AGCTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATT	2518
Db	2179	AGCTAGGCTATGCAGAAATTTCTGGAATCCTGAGAGGATAGTGGTTTATAGCAAAGTGTT	2238
Qy	2519	TGGCTTCATCTTCCAGGGGGCCCCCACTCCGTGGTGAGCCACCATCAATACAGAAAGTGA	2578
Db	2239	TAACCTTCTCTTCTACCATCTCACAC----TGTTAAGCCACTCCCAATACAAAGGGCGA	2294
Qy	2579	CCTAAGATGTACCAGCAAGATG--CCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAA	2637
Db	2295	CCTAAACAACTAGCAAAATGTTTTTCGCTTATCTCTGCGTGGATTTCATGGACTCCAAC	2354
Qy	2638	AGCCAACGT	2646
Db	2355	CCCCAAAGT	2363

RESULT 3

BX481838

LOCUS BX481838 691 bp mRNA linear EST 04-SEP-2003

DEFINITION DKFZp686M06227_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone

DKFZp686M06227 5', mRNA sequence.

ACCESSION BX481838
 VERSION BX481838.1 GI:31941164
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 691)
 AUTHORS Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
 Fobo,G., Han,M. and Wiemann,S.
 TITLE EST (Bahr,A., Lauber,J., Mewes,H.W., Weil,B., et al.)
 JOURNAL Unpublished (2003)
 COMMENT Contact: MIPS

MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No s1 sequence available.
 This clone (DKFZp686M06227) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES Location/Qualifiers
 source 1. .691
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp686M06227"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="686 (synonym: hlcc3)"
 /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"

ORIGIN

Query Match 25.5%; Score 681.8; DB 13; Length 691;
 Best Local Similarity 99.7%; Pred. No. 7.5e-147;
 Matches 683; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1742 CCGCCGCGGCCCTGCTCCCCACCTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACA 1801
 | ||| ||||||||||||||||||||||||||||||||||||||||||||
 Db 7 CGGCCACGGCCCTGCTCCCCACCTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACA 66
 Qy 1802 ACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCG 1861
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 67 ACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCG 126
 Qy 1862 CGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTTCAGT 1921
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 127 CGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTTCAGT 186
 Qy 1922 TCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATA 1981
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 187 TCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATA 246

Qy 1982 AAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCA 2041
 |||
 Db 247 AAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCA 306
 Qy 2042 TTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGA 2101
 |||
 Db 307 TTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGA 366
 Qy 2102 AACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGACCTGAGCA 2161
 |||
 Db 367 AACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGACCTGAGCA 426
 Qy 2162 GACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGAC 2221
 |||
 Db 427 GACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGAC 486
 Qy 2222 CCTACAGATGCTCAGCTACATCCGGCCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGG 2281
 |||
 Db 487 CCTACAGATGCTCAGCTACATCCGGCCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGG 546
 Qy 2282 ATGGCAGTAGAATAAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATG 2341
 |||
 Db 547 ATGGCAGTAGAATAAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATG 606
 Qy 2342 ACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGAT 2401
 |||
 Db 607 ACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGAT 666
 Qy 2402 ATGCATTTATATAGGCAACTCGATA 2426
 |||
 Db 667 ATGCATTTATATAGGCAACTCGATA 691

RESULT 4

BI330745

LOCUS BI330745 849 bp mRNA linear EST 30-JUL-2001

DEFINITION 602982409F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5135115 5', mRNA sequence.

ACCESSION BI330745

VERSION BI330745.1 GI:15015402

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 849)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLAM11332 row: a column: 04
High quality sequence stop: 758.

FEATURES Location/Qualifiers
 source 1. .849
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5135115"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Li9"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 17.3%; Score 460.6; DB 12; Length 849;
Best Local Similarity 77.3%; Pred. No. 1.7e-95;
Matches 639; Conservative 0; Mismatches 174; Indels 14; Gaps 6;

```
Qy      991 CAGCACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCT 1050
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2   CAGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCT 61

Qy     1051 GCTGACTTCTATGTGGACCTGACCAGCATTTGACAGGCGCAGCAGAGAGCAGGAATTGGCC 1110
        || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      62 GCGGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCC 121

Qy     1111 ACCAGGGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCCTGACTTAGAT 1170
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      122 ACCGTGGAGAAGGCACAGTCTCTTGCAGCCCTGTTTCTAGAAAAAGTACAAGGCTTTGAT 181

Qy     1171 GACTTTCTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGC 1230
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      182 GACTTTCTGTGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTG 241

Qy     1231 GTGACCCCACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAG 1290
        | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      242 ACCCTCACACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGATGATAGAG 298

Qy     1291 CAGTTTACGACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCCACCCTC 1350
        ||||| | || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      299 CAGTTTTCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTG 358

Qy     1351 CTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGC 1410
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      359 CTCATTTCATGGGTGCGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGC 418

Qy     1411 CATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCT 1470
        ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      419 CATGGGGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCG 478

Qy     1471 CTCATCCCTTTCAACGTCAATCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATG 1530
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      479 CTCATTCCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATG 538
```

[illegible]

RESULT 5

BF660076

LOCUS	BF660076	549 bp	mRNA	linear	EST 20-DEC-2000
-------	----------	--------	------	--------	-----------------

ACCESSION BF660076

VERSION BF660076.1 GT:11925210

SOURCE: *Mus musculus* (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 549)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Other ESTs: maa27c08.x1

Contact: Robert Strausberg, Ph.D.

Email: cqapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arranged by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

MGI:1454454

Seq primer: -40RP from Gibco
High quality sequence stop: 435.

```

FEATURES
    source          1. .549
                   /organism="Mus musculus"
                   /mol type="mRNA"

```

```

/db_xref="taxon:10090"
/clone="IMAGE:3812342"
/sex="female"
/dev_stage="10 weeks"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li10"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Library constructed by Life
Technologies."

```

ORIGIN

```

Query Match          13.9%; Score 370.4; DB 10; Length 549;
Best Local Similarity 79.7%; Pred. No. 1.1e-74;
Matches 437; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

```

Qy      1565 CTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGGAGCACTGTGCCTACATCA 1624
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2    CTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTCA 61

Qy      1625 TCATCTACGGGATGCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCCTCCAGCCCTTCC 1684
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      62   TCATCTACGCGATGCCATCTACTGGCTGACAAACCTGCGGCCCGTGCCTGAGCTCTTCC 121

Qy      1685 TGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCG 1744
          | || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      122  TTCTACACTTCCTGCTCGTGTGGTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCTG 181

Qy      1745 CCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAAC 1804
          || | ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      182  CCTCTGCCATGCTGCCACCTTCCACATGTCTCCTTCTTCTGCAATGCCCTCTACAAC 241

Qy      1805 CCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGT 1864
          ||||| ||||| | | ||||| ||||| ||||| ||||| ||||| |||||
Db      242  CCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAACCTGTGGATAGTGCCTGCAT 301

Qy      1865 GGATTTCCAAAGTGTCTTCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTCAAGTTCA 1924
          |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      302  GGATCTCCAAGCTGTCGTTCTCCGGTGGTGTCTCGGGGCTGATGCAGATTCAATTTA 361

Qy      1925 GCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAA 1984
          || || || | | | |||| |||| |||| |||| |||| |||| ||||
Db      362  ATGGACACCTTTACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGA 421

Qy      1985 TCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTG 2044
          | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      422  TGATCAGTGCCATGGACCTGAACCTCGCATCCACTCTATGCGATCTACCTCATGTGCATCG 481

Qy      2045 GCCTCAGCGGTGGCTTCATGGTCCGTGACTACGTGTCTTAAGGTTCAACAGAAAC 2104
          || |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      482  GCATCAGCTACGGCTTCCTGTTCCGTGACTATCTATCCTTGAAGCTCATCAACAGAAAGT 541

Qy      2105 CAAGTCAA 2112
          ||| ||||
Db      542  CAATTCAA 549

```

RESULT 6
 BY705076
 LOCUS BY705076 583 bp mRNA linear EST 16-DEC-2002
 DEFINITION BY705076 RIKEN full-length enriched, adult male liver *Mus musculus* cDNA clone 1300003C16 5', mRNA sequence.
 ACCESSION BY705076
 VERSION BY705076.1 GI:27116215
 KEYWORDS EST.
 SOURCE *Mus musculus* (house mouse)
 ORGANISM *Mus musculus*
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 583)
 AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 22354683
 PUBMED 12466851
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,


```

      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      308 CAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACA 367
Qy      396 GATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCAC 455
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      368 GATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCAC 427
Qy      456 TGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAG 515
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      428 AGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAG 487
Qy      516 CTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCCACGTGCGCCAGCACAAACCAGCTGCTCCC 575
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      488 TACGCCTCAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCC 547
Qy      576 CAACTTGACTGTGCGAGAGACCTTGGCCTTCATTG 610
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      548 CAACCTGACCGTCAGAGAGACCCTGGCTTTCATTG 582

```

RESULT 7

T91380/c

LOCUS T91380 457 bp mRNA linear EST 22-MAR-1995

DEFINITION yd53b02.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
IMAGE:111915 3', mRNA sequence.

ACCESSION T91380

VERSION T91380.1 GI:723293

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 457)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfling,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 827

High quality sequence stops: 379 Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 827 Std Error: 0.00

Seq primer: -21ml3

High quality sequence stop: 379.

FEATURES

source

Location/Qualifiers

1. .457

/organism="Homo sapiens"

/mol_type="mRNA"

ORIGIN

RESULT 8
BX482362

LOCUS BX482362 334 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp686F02230_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFZp686F02230 5', mRNA sequence.

ACCESSION BX482362

VERSION BX482362.1 GI:31942182

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 334)

AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.

TITLE EST (Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., et al.)

JOURNAL Unpublished (2003)

COMMENT Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Heinrich-
Heine-University, Duesseldorf/Germany) within the cDNA sequencing
consortium of the German Genome Project. No sl sequence available.
This clone (DKFZp686F02230) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source

Location/Qualifiers

1. .334

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686F02230"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="686 (synonym: hlcc3)"

/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match 12.5%; Score 332.4; DB 13; Length 334;

Best Local Similarity 99.7%; Pred. No. 5.3e-66;

Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1291 CAGTTTACGACGCTGATCCGTCGTCAGATTCCAACGACTTCCGAGACCTGCCCACCCTC 1350
|||||

Db 1 CAGTTTACGACGCTGAGCCGTCGTCAGATTCCAACGACTTCCGAGACCTGCCCACCCTC 60

Qy 1351 CTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCTCTATTTTGGC 1410
|||||

Db 61 CTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCTCTATTTTGGC 120

Qy 1411 CATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCT 1470
|||||

Db 121 CATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCT 180

Qy 1471 CTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATG 1530

```

|||||
Db      181 CTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGGCAATG 240
Qy      1531 CTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATC 1590
|||||
Db      241 CTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATC 300
Qy      1591 CTCGGGGAGCTTCCGGAGCACTGTGCCTACATCA 1624
|||||
Db      301 CTCGGGGAGCTTCCGGAGCACTGTGCCTACATCA 334

```

RESULT 9

BB610072

LOCUS BB610072 510 bp mRNA linear EST 26-OCT-2001

DEFINITION BB610072 RIKEN full-length enriched, adult male liver *Mus musculus* cDNA clone 1300007N20 5', mRNA sequence.

ACCESSION BB610072

VERSION BB610072.1 GI:16451685

KEYWORDS EST.

SOURCE *Mus musculus* (house mouse)

ORGANISM *Mus musculus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; *Mus*.

REFERENCE 1 (bases 1 to 510)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* . 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* . 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES Location/Qualifiers
 source 1. .510
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1300007N20"
 /sex="male"
 /tissue_type="liver"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult male liver"

ORIGIN

Query Match 12.4%; Score 331.2; DB 10; Length 510;
Best Local Similarity 83.7%; Pred. No. 1.3e-65;
Matches 375; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy      99 CATGGCCGGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGGCCACTCCCCAGGATAC 158
        ||||| | ||| |   ||| ||||   |||| ||| ||| |   | ||||| |
Db      63 CATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGC 122

Qy     159 CTCGGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTA 218
        ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     123 TTCGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTA 182

Qy     219 CAGTGGCCAGCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTC 278
        ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     183 CAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTC 242

Qy     279 TCAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTG 338
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     243 TCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAG 302

Qy     339 CCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGAT 398
        ||| | || ||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db     303 CCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGAT 362

Qy     399 GCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGG 458
        ||||| ||||| ||||| ||||| || ||||| || || || ||||| || ||
Db     363 GCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGG 422

Qy     459 CCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCAGCTC 518
        | |||| |||| ||||| || ||||| || || ||||| ||||| ||||| ||
Db     423 CAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTAC 482

Qy     519 GCCTCAGCTGGTGAGGAAGTGTGTGGCC 546
```

|||||
Db 483 GCCTCAGCTGGTGAGGAAGTGC GTT GCC 510

RESULT 10

AI157365

LOCUS AI157365 511 bp mRNA linear EST 30-SEP-1998

DEFINITION ui45h01.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1885393 5', mRNA sequence.

ACCESSION AI157365

VERSION AI157365.1 GI:3685834

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 511)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:969717

Seq primer: custom primer used

High quality sequence stop: 480.

FEATURES

source

Location/Qualifiers

1. .511

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="IMAGE:1885393"

/dev_stage="embryo, 14 dpc"

/lab_host="DH10B"

/clone_lib="Sugano mouse embryo mewa"

/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);

Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed

with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTTT];

double-stranded cDNA was ligated to a DraIII adaptor

[TGTTGGCCTACTGG], digested and cloned into distinct DraIII

sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site

CACCATGTG). XhoI should be used to isolate the cDNA

insert. Size selection was performed to exclude fragments

<1.5kb. Library constructed by Dr. Sumio Sugano

(University of Tokyo Institute of Medical Science).

Custom primers for sequencing: 5' end primer

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 827
High quality sequence stops: 383 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 827 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 383.

FEATURES Location/Qualifiers

source 1. .564
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:467532"
/db_xref="taxon:9606"
/clone="IMAGE:111915"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen 1NFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 11.9%; Score 318.4; DB 14; Length 564;
Best Local Similarity 98.9%; Pred. No. 1.2e-62;
Matches 352; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

Qy 1871 CCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTCAGTTCAGCAGAA 1930
|||||

Db 1 CCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTCAGTTCAGCAGAA 60

Qy 1931 GAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCA 1990
|||||

Db 61 GAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCA 120

Qy 1991 GTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCA 2050
|||||

Db 121 GTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCA 180

Qy 2051 GCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTC 2110
|||||

Db 181 GCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTC 240

Qy 2111 AAGACTGGTGATTACGCCAGACGTCTGCCCCTGGTGGGGGACCTGA-GCAGACCCTTC 2169
 |||

Db 241 AAGACTGGTGATTACGCCAGACGTCTGCCCCTGGTGGGGGACCTGAGGCAGACCCTTC 300

Qy 2170 AACTGCAC-TCCCTCCTCAGGAGCCCCCTTCCTGGGG-ACAGTGAGGACAATGACCC 2223
 |||

Db 301 AACTGCACTTCCCTCCTCAGGAGCCCCCTTCCTGGGGAACAGTGAGGACAATGAACC 356

RESULT 12

AI151811

LOCUS AI151811 500 bp mRNA linear EST 30-SEP-1998

DEFINITION ui46c10.y1 Sugano mouse embryo mewa Mus musculus cDNA clone

IMAGE:1885458 5', mRNA sequence.

ACCESSION AI151811

VERSION AI151811.1 GI:3680280

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 500)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:969782

Seq primer: custom primer used

High quality sequence stop: 499.

FEATURES

source

Location/Qualifiers

1. .500

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="IMAGE:1885458"

/dev_stage="embryo, 14 dpc"

/lab_host="DH10B"

/clone_lib="Sugano mouse embryo mewa"

/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);

Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed

with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTTT];

double-stranded cDNA was ligated to a DraIII adaptor

[TGTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

Query Match 11.6%; Score 309.8; DB 9; Length 500;
Best Local Similarity 83.0%; Pred. No. 1.2e-60;
Matches 366; Conservative 0; Mismatches 72; Indels 3; Gaps 1;

RESULT 13

LOCUS

DEFINITION vj35a03.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA clone IMAGE:930988 5', mRNA sequence.

ACCESSION AA537862

VERSION AA537862.1 GI:2283855

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 463)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:535908
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 393.

FEATURES Location/Qualifiers
source 1. .463
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:930988"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dT. Average
insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 10.4%; Score 276.8; DB 9; Length 463;
Best Local Similarity 76.3%; Pred. No. 5e-53;
Matches 354; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

Qy 1117 GAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTT 1176
||||||| ||||| || ||||||||| ||||||||| | | ||| |||||||||

Db 1 GAGAAGGCACAGTCTCTTGCAGCCCTGTTCCCTAGAAAAAGTACAAGGCTTTGATGACTTT 60

Qy 1177 CTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACC 1236
|| ||||||| ||| | ||||| || || |||| | ||| |

Db 61 CTGTGGAAAGCTGAGGCAAAGGAAGCAACACAAGCACCCACACAGTCAGCCCTGACCCTC 120

Qy 1237 CCACTAGACACCAACTGCCTCCCGAGTCCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTT 1296
||| ||||| |||| || || || || ||||| ||| | | |||||||

Db 121 ACACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTT 177

[illegible]

/db_xref="taxon:8030"
/clone_lib="gut"
/note="Vector: pBlueScriptIISK+; Library Creator: Matthew
L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
B.C.), Rachel Roper (University of Victoria)"

ORIGIN

Query Match 10.3%; Score 276; DB 14; Length 781;
Best Local Similarity 65.7%; Pred. No. 1e-52;
Matches 402; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

```
Qy      1510 TGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGT 1569
      ||| || ||||| || ||| | ||| |||| ||||| ||||| ||||| ||
Db      769 TGTCACACAGAGAGAGCTATGTTGTACCATGAGCTGGAGGACGGCATGTATAACGTCACA 710

Qy      1570 CCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCATC 1629
      | || ||||| ||||| ||||| ||||| ||||| ||||| || || | ||
Db      709 TCCTACTTCTTTGCCAAGGTCTCGGGGAGCTTCCAGAGCACTGTGTGTTACGTTGGTC 650

Qy      1630 TACGGGATGCCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTG 1689
      ||||| | |||| ||||| ||||| ||||| | | || | ||||| ||
Db      649 TACGGCCTACCCATCTACTGGCTGGCTGGCCTGAACCAGGCCCCGGACCGCTTCCTGCTC 590

Qy      1690 CACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCG 1749
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| || ||
Db      589 AACTTCCTGCTGGTGTGGCTCATGGTGTACTGCAGCCGCAGCATGGCTCTGTTTGTGGCT 530

Qy      1750 GCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTC 1809
      || | ||||| | || | | ||||| | ||||| | || | || |||
Db      529 GCAGCCTTACCCACCTGCAGACATCAGCCTTCATGGGCAATTCTCTGTTCAGTGTGTTTC 470

Qy      1810 TACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATT 1869
      ||||| | || ||||| | || | || | || ||||| ||| | ||| |
Db      469 TACCTTACTGGAGGCTTCGTCATTAGCCTGGAGAACATGTGGTTCGTGGCGTCTCGGTTTC 410

Qy      1870 TCCAAAGTGTCTTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTGAGTTCAGCAGA 1929
      ||| | | ||||| ||||| ||||| || ||||| || ||||| ||||| ||
Db      409 TCCCATGCCTCCTTCATGCGCTGGGGCTTTGAGGGCATGCTGCAGGTCCAGTTCAGGGGA 350

Qy      1930 AGAAGTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTC 1989
      | || | | |||| ||||| || || || || | || |
Db      349 CTCAAGTACCCGCTCTCCATCGGCAACTTCACCTCAACATCGATGGCATAACATGTGGTG 290

Qy      1990 AGTGCCATGGAGCTGGACTCGTACCTCTCTACGCCATCTACCTCATCGTCATTGGCCTC 2049
      || ||||| || || ||||| ||||| || ||||| | |||| | ||
Db      289 GAAGCTATGGATATGAACAGTACCTCTCTACTCCTGCTACCTGGTTCTCATCGCTGTC 230

Qy      2050 AGCGGTGGCTTCATGGTCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGT 2109
      | ||||| || || ||||| || || | ||||| ||||| || ||
Db      229 GTAGTGGGCTTCATGCTGCTCTACTACCTATCACTCAAATTCATCAAGCAGAAGTCCAGC 170

Qy      2110 CAAGACTGGTGA 2121
      || ||||| ||
Db      169 CAGGACTGGTGA 158
```

RESULT 15
 CD739823
 LOCUS CD739823 640 bp mRNA linear EST 26-JUN-2003
 DEFINITION 4028769 lGAL - Chicken Intestinal Lymphocyte Gallus gallus cDNA
 clone lGAL_21P20 5', mRNA sequence.
 ACCESSION CD739823
 VERSION CD739823.1 GI:32290672
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 640)
 AUTHORS Min,W., Lillehoj,H.S., Ashwell,C.M., Matukumalli,L.K., van
 Tassel,C. and Han,J.Y.
 TITLE Chicken intestinal lymphocyte EST database as a resource for the
 analysis of mucosal immune function
 JOURNAL Unpublished (2003)
 COMMENT Contact: Hyun S. Lillehoj
 Animal Parasite Diseases Laboratory
 Animal and Natural Resources Institute, USDA
 Bldg.1043, BARC-East, Beltsville, MD 20705, USA
 Tel: 3015048771
 Fax: 3015045103
 Email: hlilleho@anri.barc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 0.000925 using options -trim_alt '' -trim_fasta. Vector identified
 by cross_match using options -minmatch 12 -minscore 18
 Plate: 21 row: P column: 20
 Seq primer: ATTTAGGTGACACTATAG
 High quality sequence stop: 640.
 FEATURES Location/Qualifiers
 source 1. .640
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="white leghorn SC"
 /db_xref="taxon:9031"
 /clone="lGAL_21P20"
 /sex="mixed"
 /tissue_type="Gut"
 /cell_type="Lymphocyte"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="lGAL - Chicken Intestinal Lymphocyte"
 /note="Organ: Intestine; Vector: pCMV-SPORT6; Site_1:
 SalI; Site_2: NotI; Normalized library from chicken gut
 infected with coccidia duodenum and middle gut."
 ORIGIN
 Query Match 9.9%; Score 263.2; DB 14; Length 640;
 Best Local Similarity 64.7%; Pred. No. 8.5e-50;
 Matches 407; Conservative 0; Mismatches 219; Indels 3; Gaps 1;
 Qy 1351 CTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGC 1410
 || ||||| ||||| | ||||| | || || | || | |||

Db 6 CTAGTCCATGGATTTGAGGCCCTTGTGTCATGTCATTATTAATTGGATTTTTGTACTATGGC 65
 Qy 1411 CATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCT 1470
 || | ||| ||||| | || || | || ||||| |||||
 Db 66 CACGAAGGCA---GACTCTCCATTCGTGACACATCAGCACTGCTGTACATGATAGGTGCA 122
 Qy 1471 CTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATG 1530
 || ||||| |||| || ||| ||||| || | ||||| || |||||
 Db 123 CTAATCCCATTACGGTGATTTTGATGTTATTGTCTAATGTCATTGAGAAAGAGCAATG 182
 Qy 1531 CTTTACTATGAAGTGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATC 1590
 ||||| ||| ||||| | || |||| | | || || ||||| |||||
 Db 183 CTTTATCTTGACTTGGAATGGAATGTATTCTGTTACCCCGTACTTCTTTGCTAAGATT 242
 Qy 1591 CTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGG 1650
 | ||||| ||||| ||||| || | || || || || || || |||||
 Db 243 TTGGGGGAGCTTCCCAGCACTGCGCTTTCGTTATAATTTATGGGGTCCCCTACTGG 302
 Qy 1651 CTGGCCAACCTGAGGCCAGGCCCTCCAGCCCTTCTGCTGCACTTCTGCTGGTGTGGCTG 1710
 ||| | || || || | || | || ||||| ||||| | |||||
 Db 303 CTGACAAATCTATTTCTGAAGCAGAATTTTCTGCTGAAGTCTTCTCAGTGTGGCTG 362
 Qy 1711 GTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCCGGCCCTGCTCCCCACCTTCCAC 1770
 | || | ||| | | ||||| || | || || ||||| || || || ||
 Db 363 GCTGTATACTGCGCCCGTGAATGGCACTTTGGGTGGCAGCACTGCTGCCAACGTTACAG 422
 Qy 1771 ATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATG 1830
 | | | ||| | ||||| ||| | || || ||||| || || || ||
 Db 423 CTCTCAGCTTTCCTTGGCAATGTCTTTTCACTTCGTTCTACCTGAGCGGTGGTTTTGTG 482
 Qy 1831 ATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCTGCGG 1890
 |||| | || || ||||| || || || ||||| || || || || ||
 Db 483 ATAAGCCTGGAACAACCTCTGGACAGTTCCATATTGGGTTTCTAAGGTATCTTTTCTCAGA 542
 Qy 1891 TGGTGTGTTTGAAGGGCTGATGAAGATTGAGTTCAGCAGAAGAACTTATAAAATGCCTCTC 1950
 ||| ||| |||| ||||| | ||||| || | ||| ||||| |
 Db 543 TGGAATTTCCAAGGCATGATGCANATTGAGTTCAGTTCATGATGATATGCCTTTT 602
 Qy 1951 GGGAACCTCACCATCGCGGTCTCAGGAGA 1979
 ||||| |||| || | ||||| |
 Db 603 GGGAACGTCACAATTAAATTCAGGAAA 631

Search completed: February 26, 2004, 09:39:35
 Job time : 4514.21 secs

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:40:23 ; Search time 6855.55 Seconds
(without alignments)
16874.299 Million cell updates/sec

Title: US-09-989-981A-7
Perfect score: 2669
Sequence: 1 gtgtccctgctccaggaaac.....caattaaaaatgtattgagc 2669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*

```

28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%					Description
	No.	Score	Match Length DB	ID		
	1	2669	100.0	2669	6	AX685735 Sequence
	2	2655.2	99.5	2679	9	AF324494 Homo sapi
	3	2020.4	75.7	2022	9	AF320294 Homo sapi
	4	1680.6	63.0	3239	6	AX478099 Sequence
	5	1518.6	56.9	4829	10	AF351785 Rattus no
	6	1511.6	56.6	3674	10	AF324495 Mus muscu
	7	1454.2	54.5	2284	10	AY196216 Mus muscu
	8	1449.4	54.3	2285	10	AY196215 Mus muscu
	9	1430	53.6	2019	6	AX685731 Sequence
	10	724	27.1	127066	9	AC084265 Homo sapi
	11	724	27.1	182261	2	AC087053 Homo sapi
	12	722.4	27.1	139342	9	AC108476 Homo sapi
	13	663.6	24.9	2201	9	AF351824 Homo sapi
	14	581.2	21.8	185045	2	AC146466 Callithri
c	15	576.4	21.6	202533	2	AC146464 Saimiri s
	16	563	21.1	207760	2	AC146286 Callicebu
	17	558.6	20.9	178016	2	AC146787 Aotus nan
	18	399.2	15.0	159346	2	AC145533 Lemur cat
	19	274.2	10.3	660	9	AF351817 Homo sapi
	20	273.4	10.2	1292	9	AF351822 Homo sapi
c	21	260.6	9.8	68166	2	AC084712 Homo sapi
c	22	244.2	9.1	159346	2	AC145533 Lemur cat
	23	244	9.1	1323	9	AF351815 Homo sapi
c	24	231.2	8.7	250	11	G29195 human STS S
	25	230.4	8.6	237445	2	AC120701 Rattus no
c	26	230.4	8.6	312858	2	AC112747 Rattus no
c	27	229.4	8.6	40929	10	AY145899 Rattus no
	28	226.8	8.5	1387	10	AF351804 Mus muscu
c	29	226.2	8.5	204584	10	AC122243 Mus muscu
c	30	207.4	7.8	64889	2	AC120532 Oryza sat
	31	204.8	7.7	1378	10	AF351809 Mus muscu
	32	203.8	7.6	685	9	AF351820 Homo sapi
	33	203.6	7.6	1920	6	AX456519 Sequence

	34	203.6	7.6	2340	6	AX320883	AX320883 Sequence
	35	203.6	7.6	2340	6	AX685733	AX685733 Sequence
	36	203.6	7.6	2340	9	AF320293	AF320293 Homo sapi
	37	203.6	7.6	2516	6	AX456520	AX456520 Sequence
	38	203.6	7.6	2740	9	AF312715	AF312715 Homo sapi
c	39	202.4	7.6	400	11	G16507	G16507 human STS S
	40	195	7.3	2351	10	AY195873	AY195873 Mus muscu
	41	194.4	7.3	2035	6	AX456526	AX456526 Sequence
	42	194.4	7.3	2470	10	AF312714	AF312714 Rattus no
	43	194	7.3	2351	10	AY195872	AY195872 Mus muscu
	44	193.4	7.2	1915	6	AX456523	AX456523 Sequence
	45	193.4	7.2	1959	6	AX685729	AX685729 Sequence

ALIGNMENTS

RESULT 1

AX685735

LOCUS AX685735 2669 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 7 from Patent WO02081691.

ACCESSION AX685735

VERSION AX685735.1 GI:29371744

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 7 17-OCT-2002;

Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM (US)

FEATURES

Location/Qualifiers

source

1. .2669

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

CDS

100. .2121

/note="unnamed protein product; human ABCG8 (hABCG8)"

/codon_start=1

/protein_id="CAD86573.1"

/db_xref="GI:29371745"

/db_xref="REMTREMBL:CAD86573"

/translation="MAGKAAEERGLPKGATPQDTSGLQDRLFSSSEDNSLYFTYSGQP
NTLEVRDLNYQVDLASQVPWFELAQFKMPWTSPSCQNSCELGIQNLSFKVRSQGMLA
IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSPQLVRKCVAHVRQHNQLLPN
LTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSFTAHLVKTLSRLAKGNRLVLISLHQPBSDI
FRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQ
ELATREKAQSLAALFLEKVRDLDDFLWKAETKDLDEDTCESSVTPLDTNCLPSPTKM
PGAVQQFTTLIRRQISNDFRDLPTLLIHGAACLMSMTIGFLYFGHGSIQLSFMDTAA
LLFMIGALIPFNVILDVISKYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYI
IIYGMPTYWLANLRPGLQPFLHFLLVLVVFCCRIMALAAAALLPTFHMASFFSNAL
YNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAV
SGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYVSLRFIKQKPSQDW"

ORIGIN

Query Match 100.0%; Score 2669; DB 6; Length 2669;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTGTCCCTGCTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCT	60
Db	1	GTGTCCCTGCTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCT	60
Qy	61	AAGAGAGCTGCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGGAAGGCGGCAGAG	120
Db	61	AAGAGAGCTGCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGGAAGGCGGCAGAG	120
Qy	121	GAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACCTCGGGCCTCCAGGATAGATTG	180
Db	121	GAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACCTCGGGCCTCCAGGATAGATTG	180
Qy	181	TTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTG	240
Db	181	TTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTG	240
Qy	241	GAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAG	300
Db	241	GAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAG	300
Qy	301	CTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGCCAGAATTCTTGTGAGCTGGGC	360
Db	301	CTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGCCAGAATTCTTGTGAGCTGGGC	360
Qy	361	ATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCA	420
Db	361	ATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCA	420
Qy	421	GTTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATC	480
Db	421	GTTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATC	480
Qy	481	AAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAGTGT	540
Db	481	AAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAGTGT	540
Qy	541	GTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTG	600
Db	541	GTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTG	600
Qy	601	GCCTTCATTGCCCAGATGCGGCTGCCGAGAACCTTCTCCAGGCCAGCGTGACAAAAGG	660
Db	601	GCCTTCATTGCCCAGATGCGGCTGCCGAGAACCTTCTCCAGGCCAGCGTGACAAAAGG	660
Qy	661	GTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCCTGACACCCGCGTGGGCAAC	720
Db	661	GTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCCTGACACCCGCGTGGGCAAC	720
Qy	721	ATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGACAGTC	780
Db	721	ATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGACAGTC	780

Qy	781	CTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACA	840
Db	781	CTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACA	840
Qy	841	GCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATC	900
Db	841	GCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATC	900
Qy	901	TCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACG	960
Db	901	TCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACG	960
Qy	961	TCTGGCACCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATC	1020
Db	961	TCTGGCACCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATC	1020
Qy	1021	GGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATT	1080
Db	1021	GGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATT	1080
Qy	1081	GACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCC	1140
Db	1081	GACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCC	1140
Qy	1141	CTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGAT	1200
Db	1141	CTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGAT	1200
Qy	1201	CTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCACTAGACACCAACTGCCTCCCG	1260
Db	1201	CTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCACTAGACACCAACTGCCTCCCG	1260
Qy	1261	AGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATT	1320
Db	1261	AGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATT	1320
Qy	1321	TCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATG	1380
Db	1321	TCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATG	1380
Qy	1381	TCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGAT	1440
Db	1381	TCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGAT	1440
Qy	1441	ACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTC	1500
Db	1441	ACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTC	1500
Qy	1501	ATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTAC	1560
Db	1501	ATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTAC	1560
Qy	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCGGAGCACTGTGCCTAC	1620
Db	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCGGAGCACTGTGCCTAC	1620

Qy	1621	ATCATCATCTACGGGATGCCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCC	1680
Db	1621	ATCATCATCTACGGGATGCCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCC	1680
Qy	1681	TTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTG	1740
Db	1681	TTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTG	1740
Qy	1741	GCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTAC	1800
Db	1741	GCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTAC	1800
Qy	1801	AACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCC	1860
Db	1801	AACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCC	1860
Qy	1861	GCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGT'TTTGAAGGGCTGATGAAGATTGAG	1920
Db	1861	GCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGT'TTTGAAGGGCTGATGAAGATTGAG	1920
Qy	1921	TTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGAT	1980
Db	1921	TTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGAT	1980
Qy	1981	AAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTC	2040
Db	1981	AAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTC	2040
Qy	2041	ATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAG	2100
Db	2041	ATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAG	2100
Qy	2101	AAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGACCTGAGC	2160
Db	2101	AAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGACCTGAGC	2160
Qy	2161	AGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTCCCTGGGGACAGTGAGGACAATGA	2220
Db	2161	AGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTCCCTGGGGACAGTGAGGACAATGA	2220
Qy	2221	CCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAG	2280
Db	2221	CCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAG	2280
Qy	2281	GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTCTGCTCACTGGCAGGAGACTGCGAT	2340
Db	2281	GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTCTGCTCACTGGCAGGAGACTGCGAT	2340
Qy	2341	GACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGA	2400
Db	2341	GACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGA	2400
Qy	2401	TATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAAC TAGGAATGAATTGGGTAG	2460
Db	2401	TATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAAC TAGGAATGAATTGGGTAG	2460
Qy	2461	CTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTG	2520

```

|||||
Db      2461 CTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTG 2520
Qy      2521 GCTTCATCTTCCAGGGGGCCCCCACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACC 2580
|||||
Db      2521 GCTTCATCTTCCAGGGGGCCCCCACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACC 2580
Qy      2581 TAAGATGTACCAGCAAGATGCCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAAAGC 2640
|||||
Db      2581 TAAGATGTACCAGCAAGATGCCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAAAGC 2640
Qy      2641 CAACGTGAACAATTAATAATGTATTGAGC 2669
|||||
Db      2641 CAACGTGAACAATTAATAATGTATTGAGC 2669

```

RESULT 2

AF324494

LOCUS AF324494 2679 bp mRNA linear PRI 07-AUG-2001

DEFINITION Homo sapiens sterolin-2 (ABCG8) mRNA, complete cds.

ACCESSION AF324494

VERSION AF324494.1 GI:15088539

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2679)

AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H.,
Ose,L., Stalenhoef,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E.,
Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and
Patel,S.B.

TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterolin-1 and
sterolin-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)

MEDLINE 21344600

PUBMED 11452359

REFERENCE 2 (bases 1 to 2679)

AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
Street, STB541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers

source

1. .2679

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="2"

/map="2p21; between D2S2294 and D2S2298"

/tissue_type="liver"

gene

1. .2679

/gene="ABCG8"

CDS

91. .2112

/gene="ABCG8"

/codon_start=1

/product="sterolin-2"
/protein_id="AAK84078.1"
/db_xref="GI:15088540"
/translation="MAGKAAEERGLPKGATPQDTSGLQDRLFSSSDNSLYFTYSGQP
NTLEVRDLNCQVDLASQVPWFQLAQFKMPWTS PSCQNSCELGIQNLSFKVRSQMLA
IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSPQLVRKCVAHVRQHNQLLPN
LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER
RRVSIQVQLLWNPGLILDEPTSGLDSFTAHLVKTLSRLAKGNRLVLI SLHQPRSDI
FRLFDLVLLMTSGTPIYLGAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQ
ELATREKAQSLAALFLEKVRDLDDFLWKAETKDLEDTCVESSVTPLDNCPLSPPTKM
PGAVQQFTTLIRRQISNDFRDLPTLLIHGAELMSMTIGFLYFGHGSIQLSFMDTAA
LLFMIGALIPFNVILDVISKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYI
IIYGMPTYWLANLRPGLQPFLLHFLLVLVVFCCRIMALAAAALLPTFHMASFFSNAL
YNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEGLMKIQFSRRTYKMPGLNLTIAV
SGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYVSLRFIKQKPSQDW"

ORIGIN

Query Match 99.5%; Score 2655.2; DB 9; Length 2679;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2657; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	10	CTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCTAAGAGAGCT	69
Db	1	CTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCTAAGAGAGCT	60
Qy	70	GCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGAAGGCGGCAGAGGAGAGAGGG	129
Db	61	GCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGAAGGCGGCAGAGGAGAGAGGG	120
Qy	130	CTGCCGAAAGGGGCCACTCCCCAGGATACCTCGGGCCTCCAGGATAGATTGTTCTCCTCT	189
Db	121	CTGCCGAAAGGGGCCACTCCCCAGGATACCTCGGGCCTCCAGGATAGATTGTTCTCCTCT	180
Qy	190	GAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTGGAGGTCAGA	249
Db	181	GAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTGGAGGTCAGA	240
Qy	250	GACCTCAACTACCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAGCTGGCTCAG	309
Db	241	GACCTCAACTGCCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAGCTGGCTCAG	300
Qy	310	TTCAAGATGCCCTGGACATCTCCAGCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAAC	369
Db	301	TTCAAGATGCCCTGGACATCTCCAGCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAAC	360
Qy	370	CTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGG	429
Db	361	CTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGG	420
Qy	430	AGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGC	489
Db	421	AGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGC	480
Qy	490	CAGATCTGGATCAATGGGCAGCCAGCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCAC	549
Db	481	CAGATCTGGATCAATGGGCAGCCAGCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCAC	540

Qy	550	GTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTGGCCTTCATT	609
Db	541	GTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTGGCCTTCATT	600
Qy	610	GCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCGTGACAAAAGGGTGGAGGAC	669
Db	601	GCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCGTGACAAAAGGGTGGAGGAC	660
Qy	670	GTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTACGTG	729
Db	661	GTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTACGTG	720
Qy	730	CGGGGGTTGTCTGGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAAC	789
Db	721	CGGGGGTTGTCTGGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAAC	780
Qy	790	CCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGCCCACAAC	849
Db	781	CCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGCCCACAAC	840
Qy	850	CTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCAC	909
Db	841	CTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCAC	900
Qy	910	CAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCACC	969
Db	901	CAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCACC	960
Qy	970	CCCATCTACTTAGGGGCGGGCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACCCC	1029
Db	961	CCCATCTACTTAGGGGCGGGCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACCCC	1020
Qy	1030	TGTCCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATTGACAGGCGC	1089
Db	1021	TGTCCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATTGACAGGCGC	1080
Qy	1090	AGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTA	1149
Db	1081	AGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTA	1140
Qy	1150	GAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGATCTTGACGAG	1209
Db	1141	GAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGATCTTGACGAG	1200
Qy	1210	GACACCTGTGTGGAAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCGAGTCCTACG	1269
Db	1201	GACACCTGTGTGGAAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCGAGTCCTACG	1260
Qy	1270	AAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTGAGTTTCCAACGAC	1329
Db	1261	AAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTGAGTTTCCAACGAC	1320
Qy	1330	TTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACC	1389
Db	1321	TTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACC	1380
Qy	1390	ATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCC	1449

Db	1381	 ATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCC	1440
Qy	1450	CTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAA	1509
Db	1441	 CTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAA	1500
Qy	1510	TGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGT	1569
Db	1501	 TGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGT	1560
Qy	1570	CCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCATC	1629
Db	1561	 CCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCATC	1620
Qy	1630	TACGGGATGCCCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCCTCCAGCCCTTCCTGCTG	1689
Db	1621	 TACGGGATGCCCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCCTCCAGCCCTTCCTGCTG	1680
Qy	1690	CACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCG	1749
Db	1681	 CACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCG	1740
Qy	1750	GCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCTTC	1809
Db	1741	 GCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCTTC	1800
Qy	1810	TACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATT	1869
Db	1801	 TACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATT	1860
Qy	1870	TCCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTAGTTTACGAGAG	1929
Db	1861	 TCCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTAGTTTACGAGAG	1920
Qy	1930	AGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTC	1989
Db	1921	 AGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTC	1980
Qy	1990	AGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTC	2049
Db	1981	 AGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTC	2040
Qy	2050	AGCGGTGGCTTCATGGTCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGT	2109
Db	2041	 AGCGGTGGCTTCATGGTCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGT	2100
Qy	2110	CAAGACTGGTGATTACAGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGCAGACCCTTC	2169
Db	2101	 CAAGACTGGTGATTACAGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGCAGACCCTTC	2160
Qy	2170	AACTGCACTCCCTCCTCAGGAGCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGA	2229
Db	2161	 AACTGCACTCCCTCCTCAGGAGCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGA	2220
Qy	2230	TGCTCAGCTACATCCGGCCCAGGGTGTGCACTGGCACAGACCAGCCACAGGATGGCAGT	2289

Db 2221 TGCTCAGCTACATCCGGCCCCAGGGTGCTGCGGTGGCACAGACCAGCCACAGGATGGCAGT 2280

Qy 2290 AGAATAAAGACAGTTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAG 2349
 |||

Db 2281 AGAATAAAGACAGTTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAG 2340

Qy 2350 AAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTT 2409
 |||

Db 2341 AAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTT 2400

Qy 2410 ATATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGT 2469
 |||

Db 2401 ATATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGT 2460

Qy 2470 GCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCT 2529
 |||

Db 2461 GCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTACCTAGCAGATTTGGCTTCATCT 2520

Qy 2530 TCCAGGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTA 2589
 |||

Db 2521 TCCAGGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTA 2580

Qy 2590 CCAGCAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAA 2649
 |||

Db 2581 CCAGCAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAA 2640

Qy 2650 CAATTAAAAATGTATTGAGC 2669
 |||

Db 2641 CAATTAAAAATGTATTGAGC 2660

RESULT 3

AF320294

LOCUS AF320294 2022 bp mRNA linear PRI 13-DEC-2000

DEFINITION Homo sapiens ABCG8 (ABCG8) mRNA, complete cds.

ACCESSION AF320294

VERSION AF320294.1 GI:11692801

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2022)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J., Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Accumulation of Dietary Cholesterol in Sitosterolemia Caused by Mutations in Adjacent ABC Transporters

JOURNAL Science (2001) In press

REFERENCE 2 (bases 1 to 2022)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J., Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Direct Submission

JOURNAL Submitted (09-NOV-2000) Molecular Genetics, University of Texas, Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9046, USA

FEATURES Location/Qualifiers

source 1. .2022

gene /organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. .2022
/gene="ABCG8"
CDS 1. .2022
/gene="ABCG8"
/note="ATP-binding cassette, subfamily G, member 8"
/codon_start=1
/product="ABCG8"
/protein_id="AAG40004.1"
/db_xref="GI:11692802"
/translation="MAGKAAEERGLPKGATPQDTSGLQDRLFSSSEDNSLYFTYSGQP
NTLEVRDLNYQVDLASQVPWFEOQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQGMLA
IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSPQLVRKCVAHVRQHNQLLPN
LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTL SRLAKGNRLVLSLHQPRSDI
FRLFDLVLLMTSGTPIYLGAAQHMQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQ
ELATREKAQSLAALFLEKVRDLDDFLWKAETKDLEDTCVESSVTPLDTNCLPSPTKM
PGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACLSMTIGFLYFGHSGSIQLSFMDTAA
LLFMIGALIPFNVILDVISKYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYI
IIYGMPTYWLANLRPGLQPFLLHFLLVLVFCCRIMALAAAALLPTFHMAFFSNAL
YNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEGLMKIQFSRRTYKMPGLNLTIAV
SGDKILSVMELDSYPLYAIYLIVIGLSGGFMVLYVSLRFIKQKPSQDW"

ORIGIN

Query Match 75.7%; Score 2020.4; DB 9; Length 2022;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2021; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	100	ATGGCCGGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACC	159
Db	1	ATGGCCGGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACC	60
Qy	160	TCGGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTAC	219
Db	61	TCGGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTAC	120
Qy	220	AGTGGCCAGCCCAACACCCTGGAGGTGAGAGACCTCAACTACCAGGTGGACCTGGCCTCT	279
Db	121	AGTGGCCAGCCCAACACCCTGGAGGTGAGAGACCTCAACTACCAGGTGGACCTGGCCTCT	180
Qy	280	CAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTGC	339
Db	181	CAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTGC	240
Qy	340	CAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATG	399
Db	241	CAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATG	300
Qy	400	CTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGC	459
Db	301	CTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGC	360
Qy	460	CGAGGTACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCAGCTCG	519
Db	361	CGAGGTACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCAGCTCG	420

Qy	520	CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAAC	579
Db	421	CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAAC	480
Qy	580	TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCC	639
Db	481	TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCC	540
Qy	640	CAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGC	699
Db	541	CAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGC	600
Qy	700	GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTCGGGGGGTGAGCGCAGGAGA	759
Db	601	GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTCGGGGGGTGAGCGCAGGAGA	660
Qy	760	GTCAGCATTGGGGTGACGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC	819
Db	661	GTCAGCATTGGGGTGACGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC	720
Qy	820	TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAA	879
Db	721	TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAA	780
Qy	880	GGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTT	939
Db	781	GGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTT	840
Qy	940	GATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACATG	999
Db	841	GATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACATG	900
Qy	1000	GTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTC	1059
Db	901	GTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTC	960
Qy	1060	TATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAG	1119
Db	961	TATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAG	1020
Qy	1120	AAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCCTGACTTAGATGACTTTCTA	1179
Db	1021	AAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCCTGACTTAGATGACTTTCTA	1080
Qy	1180	TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCA	1239
Db	1081	TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCA	1140
Qy	1240	CTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG	1299
Db	1141	CTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG	1200
Qy	1300	ACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCACCCCTCCTCATCCAT	1359
Db	1201	ACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCACCCCTCCTCATCCAT	1260

Qy	1360	GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGC	1419
Db	1261	GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGC	1320
Qy	1420	ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCT	1479
Db	1321	ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCT	1380
Qy	1480	TTCAACGTCATTCTGGATGTCTCTCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT	1539
Db	1381	TTCAACGTCATTCTGGATGTCTCTCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT	1440
Qy	1540	GAAGTGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG	1599
Db	1441	GAAGTGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG	1500
Qy	1600	CTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAAC	1659
Db	1501	CTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAAC	1560
Qy	1660	CTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC	1719
Db	1561	CTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC	1620
Qy	1720	TGTTGCAGGATTATGGCCCTGGCCGCCCGGGCCCTGCTCCCCACCTTCCACATGGCCTCC	1779
Db	1621	TGTTGCAGGATTATGGCCCTGGCCGCCCGGGCCCTGCTCCCCACCTTCCACATGGCCTCC	1680
Qy	1780	TTCTTCAGCAATGCCCTCTACAACCTCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG	1839
Db	1681	TTCTTCAGCAATGCCCTCTACAACCTCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG	1740
Qy	1840	AGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTT	1899
Db	1741	AGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTT	1800
Qy	1900	GAAGGGCTGATGAAGATTTCAGTTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTC	1959
Db	1801	GAAGGGCTGATGAAGATTTCAGTTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTC	1860
Qy	1960	ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTC	2019
Db	1861	ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGTCATGGAGCTGGACTCGTACCCTCTC	1920
Qy	2020	TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCGTACTACGTG	2079
Db	1921	TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCGTACTACGTG	1980
Qy	2080	TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA	2121
Db	1981	TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA	2022

RESULT 4

AX478099

LOCUS AX478099 3239 bp DNA linear PAT 12-AUG-2002

DEFINITION Sequence 29 from Patent WO0240541.

ACCESSION AX478099
 VERSION AX478099.1 GI:22217059
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Tang,Y.T., Yue,H., Nguyen,D.B., Hafalia,A.J., Elliott,V.S., Lu,Y.,
 Walia,N.K., Yao,M.G., Baughn,M.R., Gandhi,A.R., Ding,L.,
 Sanjanwala,M., Ramkumar,J., Arvizu,C., Gietzen,K.J., Lal,P.G.,
 Azimzai,Y., Khan,F.A., Thangavelu,K., Thornton,M., Lu,D.A.,
 Tribouley,C.M., Warren,B.A., Ison,C.H., Das,D., Raumann,B.E.,
 Policky,J.L. and Kearney,L.
 TITLE Transporters and ion channels
 JOURNAL Patent: WO 0240541-A 29 23-MAY-2002;
 Incyte Genomics, Inc. (US)
 FEATURES Location/Qualifiers
 source 1. .3239
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 6585710CB1"

ORIGIN

Query Match 63.0%; Score 1680.6; DB 6; Length 3239;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1683; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	983	GGGCGGCCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACA	1042
Db	12	GGGCGGCCCAGCACATGGTCCATTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACA	71
Qy	1043	GCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGG	1102
Db	72	GCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGG	131
Qy	1103	AATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTG	1162
Db	132	AATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTG	191
Qy	1163	ACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGG	1222
Db	192	ACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGG	251
Qy	1223	AAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGG	1282
Db	252	AAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGG	311
Qy	1283	CGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGC	1342
Db	312	CGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGC	371
Qy	1343	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	1402
Db	372	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	431

Qy	1403	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA	1462
Db	432	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA	491
Qy	1463	TCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGA	1522
Db	492	TCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGA	551
Qy	1523	GGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG	1582
Db	552	GGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG	611
Qy	1583	CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCCA	1642
Db	612	CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCCA	671
Qy	1643	CCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACCTCCTGCTGG	1702
Db	672	CCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACCTCCTGCTGG	731
Qy	1703	TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA	1762
Db	732	TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA	791
Qy	1763	CCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGG	1822
Db	792	CCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGG	851
Qy	1823	GCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCCT	1882
Db	852	GCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCCT	911
Qy	1883	TCCTGCGGTGGTGTTTTGAAGGGCTGATGAAGATTTCAGTTCAGCAGAAGAACCTTATAAAA	1942
Db	912	TCCTGCGGTGGTGTTTTGAAGGGCTGATGAAGATTTCAGTTCAGCAGAAGAACCTTATAAAA	971
Qy	1943	TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC	2002
Db	972	TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC	1031
Qy	2003	TGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA	2062
Db	1032	TGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA	1091
Qy	2063	TGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGAT	2122
Db	1092	TGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGAT	1151
Qy	2123	TCACGCCAGACGTCTGCCCCTGGTGGGGGACCTGAGCAGACCCTTCAACTGCACTCCCT	2182
Db	1152	TCACGCCAGACGTCTGCCCCTGGTGGGGGACCTGAGCAGACCCTTCAACTGCACTCCCT	1211
Qy	2183	CCTCAGGAGCCCCCTTCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACAT	2242
Db	1212	CCTCAGGAGCCCCCTTCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACAT	1271
Qy	2243	CCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAG	2302

```

      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1272 CCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAG 1331
Qy      2303 TCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTC 2362
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1332 TCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTC 1391
Qy      2363 GGTGGCACCTACAACGTTGCTAATTTATTTTCCTTTTGATATGCATTTATATAGGCAACTC 2422
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1392 GGTGGCACCTACAACGTTGCTAATTTATTTTCCTTTTGATATGCATTTATATAGGCAACTC 1451
Qy      2423 GATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTG 2482
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1452 GATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTG 1511
Qy      2483 GAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCAGGGGCCCCA 2542
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1512 GAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCAGGGGCCCCA 1571
Qy      2543 CACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATGCC 2602
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1572 CACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATGCC 1631
Qy      2603 ATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAATTAAAAATGT 2662
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1632 ATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAATTAAAAATGT 1691
Qy      2663 ATTGAGC 2669
      |||||||
Db      1692 ATTGAGC 1698

```

RESULT 5
AF351785

```

LOCUS      AF351785                      4829 bp    mRNA    linear    ROD 26-AUG-2002
DEFINITION Rattus norvegicus sterolin-2 (Abcg8) mRNA, complete cds.
ACCESSION  AF351785
VERSION    AF351785.2  GI:22477145
KEYWORDS   .
SOURCE     Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 4829)
  AUTHORS  Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H.,
            Ose,L., Stalenhoef,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E.,
            Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and
            Patel,S.B.
  TITLE    Two genes that map to the STSL locus cause sitosterolemia: genomic
            structure and spectrum of mutations involving sterolin-1 and
            sterolin-2, encoded by ABCG5 and ABCG8, respectively
  JOURNAL  Am. J. Hum. Genet. 69 (2), 278-290 (2001)
  MEDLINE  21344600
  PUBMED   11452359
REFERENCE  2 (bases 1 to 4829)
  AUTHORS  Lu,K., Yu,H., Lee,M. and Patel,S.B.

```


Qy	219	CAGTGGCCAGCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTC	278
Db	230	CAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATGGCCTC	289
Qy	279	TCAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTG	338
Db	290	TCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGTTACCGTGGAGGTCTCGCGGCAG	349
Qy	339	CCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGAT	398
Db	350	CCAGGACTCCTGGGATCTGGGCATCCGAAATCTGAGCTTCAAAGTGAGGAGTGGACAGAT	409
Qy	399	GCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGG	458
Db	410	GCTGGCTATCATAGGGAGCGCAGGCTGCGGGAGAGCCACATTACTCGACGTTATCACAGG	469
Qy	459	CCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTC	518
Db	470	CAGAGACCATGGTGGCAAGATGAAATCAGGACAAATCTGGATAAACGGGCAACCCAGCAC	529
Qy	519	GCCTCAGCTGGTGAGGAAGTGTGTGGCCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAA	578
Db	530	GCCTCAGCTGATACAGAAGTGTGTGGCACATGTGCGCCAGCAAGACCAGCTGCTCCCCAA	589
Qy	579	CTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGGCTGCCCAGAACCTTCTC	638
Db	590	TCTGACTGTCAGAGAGACCCTGACTTTCATCGCCCAGATGCGCCTGCCAAGACCTTCTC	649
Qy	639	CCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTG	698
Db	650	TCAGGCCCAGCGAGACAAACGGGTGGAAGACGTGATTGCGGAGCTGCGGCTGCGGCAGTG	709
Qy	699	CGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAG	758
Db	710	CGCCAACACCCGCGTGGGCAACACATACGTACGCGGGGTGTCCGGGGGCGAGCGCCGAAG	769
Qy	759	AGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCAC	818
Db	770	AGTGAGCATCGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATCTCTGGATGAACCCAC	829
Qy	819	CTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAA	878
Db	830	TTCCGGCCTCGACAGCTTCACCGCTCACAACCTGGTGAGAACTTTGTCCCGCCTGGCCAA	889
Qy	879	AGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTT	938
Db	890	AGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATT	949
Qy	939	TGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACAT	998
Db	950	TGACCTGGTCCTTCTGATGACGTCTGGCACCCCTATCTACCTGGGGGTGGCACAGCACAT	1009
Qy	999	GGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTT	1058
Db	1010	GGTGCAGTACTTTACATCAATTGGCTACCCTTGTCCTCGCTACAGCAACCTGCTGACTT	1069
Qy	1059	CTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGA	1118

Db	1070	CTACGTGGACTTGACGAGCATTGACAGGCGCAGCAAAGAACAGGAGGTGGCCACCATGGA	1129
Qy	1119	GAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTTCT	1178
Db	1130	GAAGGCTCGATTACTTGACGCCTTGTTCTAGAAAAAGTGCAAGGCTTTGACGACTTTTCT	1189
Qy	1179	ATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCC	1238
Db	1190	GTGGAAAGCTGAGGCAAAGAGTCTCGACACAGGCACCTATGCAGTCAGCCAGACCCTCAC	1249
Qy	1239	ACTAGACACCAACTGCCTCCCGAGTCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTAC	1298
Db	1250	ACAGGACACCAACTG---TGGAAGTGTGCTGAGCTGCCCGGGATGATACAGCAGTTTAC	1306
Qy	1299	GACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCCCACCTCCTCATCCA	1358
Db	1307	CACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACTTGCCCCACCTGTTTCATCCA	1366
Qy	1359	TGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAG	1418
Db	1367	TGGAGCAGAAGCCTGCCTGATGTCTCTCATCATTGGCTTCCTTTACTACGGCCACGCAGA	1426
Qy	1419	CATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCC	1478
Db	1427	CAAGCCGCTCTCCTTCATGGACATGGCAGCCCTCCTGTTTCATGATAGGAGCACTCATTC	1486
Qy	1479	TTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTA	1538
Db	1487	TTTTAATGTCATTCTGGATGTCGTCTCCAAATGTCACTCGGAGCGGTGCTGCTGTACTA	1546
Qy	1539	TGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGA	1598
Db	1547	TGAACTGGAGGACGGACTGTACACTGCTGGTCCTTATTTCTTTGCCAAGGTCTCGGTGA	1606
Qy	1599	GCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCCACCTACTGGCTGGCCAA	1658
Db	1607	GCTGCCAGAGCACTGTGCCTATGTCATCATCTATGGGATGCCCATCTACTGGCTGACCAA	1666
Qy	1659	CCTGAGGCCAGGCCTCCAGCCCTTCTGCTGCACCTCCTGCTGGTGTGGCTGGTGGTCTT	1718
Db	1667	CCTGCGGCCAGGGCCTGAGCTCTTCTCCTGCACTTCATGCTTCTGTGGCTGGTGGTGT	1726
Qy	1719	CTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTC	1778
Db	1727	CTGCTGCAGGACCATGGCCCTGGCCGCCTCTGCCATGCTGCCACCTTCCACATGTCTC	1786
Qy	1779	CTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTT	1838
Db	1787	CTTCTGCTGCAACGCTCTCTACAACCTCCTTCTACCTTACGGCTGGCTTCATGATAAACTT	1846
Qy	1839	GAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGT	1898
Db	1847	GAACAACCTGTGGATAGTACCTGCATGGATTTCCAAGATGTCGTTCCTCCGGTGGTGT	1906
Qy	1899	TGAAGGGCTGATGAAGATTGAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCT	1958

Db 1907 CTCAGGGCTGATGCAGATTCAGTTTAAATGGACACATTTACACCACGCAGATCGGCAACCT 1966

Qy 1959 CACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCT 2018
 |||| || | ||| | |||| | | ||| ||||| || | ||| | ||||

Db 1967 CACCTTCTCCGTCCCCGGAGACGCGATGGTCACTGCCATGGACCTGAACCTCACATCCTCT 2026

Qy 2019 CTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGT 2078
 || || ||||| ||||| ||||| ||||| || ||||| || |

Db 2027 TTATGCGATCTACCTCATCGTCATTGGCATCAGCTGTGGCTTCCTGTCCCTGTATTATCT 2086

Qy 2079 GTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGATTCACGCCAGACGTCTG 2138
 ||||| | ||||| ||||| ||| ||||| ||||| ||| | |

Db 2087 GTCCTTGAAGTTTCATCAAACAGAAAGTCAATTCAAGATTGGTGAT---GTTACAGCCTTGCT 2143

Qy 2139 CCCGCTGGTGGGGGACCTGAGCAGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCTTC 2198
 || ||||| | | || || || || ||||| |||||

Db 2144 TCCACTGGTGGGACCTTCTGCCTGGGCT-----GGCCGCCTCCTGAGGAGCCC---- 2192

Qy 2199 CTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTG 2258
 | ||||| ||||| || ||||| ||| |||| | | |||||

Db 2193 -----GACTGAGGACAATGATCCACAGATCTCAAGCAGCATCGGCGTCTTGGTGCTG 2245

Qy 2259 CAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAGTCGAAAGGGATTTCTG 2318
 ||||| ||||| ||||| ||||| ||||| ||||| || ||| |||||

Db 2246 CAGTGGCACAGGTCAGCCACAGGATGGCAGTAGAATAAAGACAGTTGAGAGG--TTTCTG 2303

Qy 2319 CTCCTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACG 2378
 |||| ||| | | || | || | | || | | |||| ||||

Db 2304 CTCACAGGCCTGGGCTTGTG-----AAACAGGTACTTCGTGAACCTGTAACG 2350

Qy 2379 TTGCTAATTTATTTCTTTTGATATGCATTTATATAGGCAACTCGATATAGGATGGGAGC 2438
 |||| ||| |||| ||| ||||| |||| ||| ||| ||||| |

Db 2351 TTGCTCATTCAATTT-----TATATCTCTATATAAACAACCCAGTATGGAATGGGAAC 2402

Qy 2439 AAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACA 2498
 || || ||||| ||||| || ||||| | | ||||| ||| | | |

Db 2403 CAATTATTTATGAATTGAGTAGCTAGGCTATGCAGAGACTTGTGGAACCCCGAGAGGATA 2462

Qy 2499 ATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCAGGGGCCCCACACTCCGTGGTGAGCC 2558
 | | | |||| |||| || ||| | | ||| | | ||||

Db 2463 GTGGTCTGTAGCAAAACATTTAGCTTTCTCCACCA-----ATCTACCCTGTTAAGCC 2515

Qy 2559 ACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATGCCATCCCTTCTTTTGTG 2618
 | | ||||| ||||| | || || | || | || | ||||

Db 2516 GCTCCCGATACAGAGGGTGACCTAAACTGACTAG--AAATGTCCTCTCTTATCTCTGTG 2573

Qy 2619 TGGGGTCATGGGCTCCAAAAGCC 2641
 ||| |||| || | || |

Db 2574 TGGCTCCATGGACTTCCAGAGTC 2596

RESULT 6
 AF324495

LOCUS AF324495 3674 bp mRNA linear ROD 07-AUG-2001
 DEFINITION Mus musculus sterolin-2 (Abcg8) mRNA, complete cds.
 ACCESSION AF324495
 VERSION AF324495.1 GI:15088541

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3674)

AUTHORS Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H.,
Ose, L., Stalenhoeft, A.F., Mietinnen, T., Bjorkhem, I., Bruckert, E.,
Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Srivastava, A. and
Patel, S.B.

TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterolin-1 and
sterolin-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)

MEDLINE 21344600

PUBMED 11452359

REFERENCE 2 (bases 1 to 3674)

AUTHORS Lu, K., Lee, M.-H. and Patel, S.B.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
Street, STB541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers

source 1. .3674
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/tissue_type="liver"

gene 1. .3674
/gene="Abcg8"

CDS 102. .2123
/gene="Abcg8"
/note="ABCG8"
/codon_start=1
/product="sterolin-2"
/protein_id="AAK84079.1"
/db_xref="GI:15088542"
/translation="MAEKTKEETQLWNGTVLQDASQGLQDSLFSSSESDNSLYFTYSQG
SNTLEVRDLTYQVDIASQVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQML
AIIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLP
NLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTRYVRGVS
RRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSD
IFRLFDLVLLMTSGTPIYLGAAQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKE
REVATVEKAQSLAALFLEKVQGFDDFLWKAELAKELNTSTHTVSLTLTQDTDCGTAVEL
PGMIEQFSTLIRRIQISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAA
LLFMIGALIPFNVILDVVSCHSERSMLYIELEDGLYTAGPYFFAKILGELPEHCAYV
IIYAMPIYWLTNLRPVPELFLHFLVWLVVFCRTMALAASAMLPFHMSSFFCNAL
YNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSI
LGDTMISAMDLSHPLYAIYLVIVIGISYGFLFLYYLSLKLKQKSIQDW"

ORIGIN

Query Match 56.6%; Score 1511.6; DB 10; Length 3674;
Best Local Similarity 77.0%; Pred. No. 0;
Matches 1965; Conservative 0; Mismatches 534; Indels 53; Gaps 8;

Qy 99 CATGGCCGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATAC 158

Db	101	CATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGC	160
Qy	159	CTC---GGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCAC	215
Db	161	TTCGCAGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCAC	220
Qy	216	CTACAGTGGCCAGCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGC	275
Db	221	CTACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGC	280
Qy	276	CTCTCAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAGATGCCCTGGACATCTCCCAG	335
Db	281	CTCTCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAGATACCCTGGAGGTCTCATAG	340
Qy	336	CTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCA	395
Db	341	CAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACA	400
Qy	396	GATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCAC	455
Db	401	GATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCAC	460
Qy	456	TGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAG	515
Db	461	AGGCAGAGGCCACGGTGGAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAG	520
Qy	516	CTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCC	575
Db	521	TACGCCTCAGCTGGTGAGGAAGTGCCTTGCATGTGCGGCAGCATGACCAACTGCTGCC	580
Qy	576	CAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTT	635
Db	581	CAACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTT	640
Qy	636	CTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCA	695
Db	641	CTCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCA	700
Qy	696	GTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGGTGAGCGCAG	755
Db	701	GTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCG	760
Qy	756	GAGAGTCAGCATTGGGGTGAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACC	815
Db	761	ACGAGTGAGCATTGGGGTGAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACC	820
Qy	816	CACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGC	875
Db	821	CATTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCTGGC	880
Qy	876	CAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCT	935
Db	881	CAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCT	940
Qy	936	GTTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCA	995

Db	941	ATTTGACCTGGTCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCAGCA	1000
Qy	996	CATGGTCCAGTATTTTCACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGA	1055
Db	1001	AATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGA	1060
Qy	1056	CTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAG	1115
Db	1061	CTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGT	1120
Qy	1116	GGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTT	1175
Db	1121	GGAGAAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTT	1180
Qy	1176	TCTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGAC	1235
Db	1181	TCTGTGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTGACCCT	1240
Qy	1236	CCCACTAGACACCAACTGCCTCCCAGTCCCTACGAAGATGCCTGGGGCGGTGCAGCAGTT	1295
Db	1241	CACACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTT	1297
Qy	1296	TACGACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCCACCCTCCTCAT	1355
Db	1298	TTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCAT	1357
Qy	1356	CCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGG	1415
Db	1358	TCATGGGTGCGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGG	1417
Qy	1416	GAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCAT	1475
Db	1418	GGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCAT	1477
Qy	1476	CCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTA	1535
Db	1478	TCCTTTCAATGTCATCCTGGATGTCGTCCTCCAAATGTCACTCGGAGAGGTCAATGCTGTA	1537
Qy	1536	CTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGG	1595
Db	1538	CTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGG	1597
Qy	1596	GGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGC	1655
Db	1598	AGAATTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGAC	1657
Qy	1656	CAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGT	1715
Db	1658	AAACCTGCGGCGCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGT	1717
Qy	1716	CTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGC	1775
Db	1718	CTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTC	1777
Qy	1776	CTCCTTCTTCAGCAATGCCCTCTACAACCTCTTCTACCTCGCCGGGGGCTTCATGATAAA	1835
Db	1778	CTCCTTCTTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAA	1837

Qy 1836 CTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTG 1895
 |||| | | ||||| ||||| | | |||| | | | | ||||| |||||
 Db 1838 CTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTG 1897

Qy 1896 TTTTGAAGGGCTGATGAAGATTTCAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAA 1955
 || | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1898 CTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAA 1957

Qy 1956 CCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCC 2015
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1958 CTTCACTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCC 2017

Qy 2016 TCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCTGTACTA 2075
 |||| | | ||||| ||||| || | |||| | | | | |||| | | |||||
 Db 2018 ACTCTATGCGATCTACCTCATTGTATCGGCATCAGCTACGGCTTCCTGTTCTGTACTA 2077

Qy 2076 CGTGTCTTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGATTACAGCCAGACGT 2135
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 2078 TCTATCCTTGAAGCTCATCAAACAGAAAGTCAATTCAAGACTGGTGATACTCAGCCTTGCT 2137

Qy 2136 CTGCCCCGCTGGTGGGGGACCTGAGCAGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCC 2195
 || | | || | | | | | | | | | | | | | | | | | | | | |
 Db 2138 CTCCTGCGCG-----GACCCTTTTCCCGGGGCTGGCCACCCAGGAGAGCC 2185

Qy 2196 TTCCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACATCCGGCCAGGGTG 2255
 ||||| | | || | | | || | | | | | | | | | | | | | |
 Db 2186 GGAAGTGGGGACAAGGCTCACACAGATCTCTCAG-----GCAGCAGCCACCTCTTAGTG 2238

Qy 2256 CTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAGTCGAAAGGGATTT 2315
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2239 CTGCAGTGGCACAGGTGAGCCACAGGATGGCAGTAGAATAAAGACAGTTGAGAGGTGTTT 2298

Qy 2316 CTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTCGGTGGCACCTACA 2375
 ||||| | || | | | || | | | | | | | | | | | | | | | |
 Db 2299 CTGCTCCCAGGCCAGGCTTGTGATGGGAGAGAGAGAA-----ACCAGGT 2343

Qy 2376 ACGTTGCTAATTTATTTCTTTTGATATGCATTTATATAGGCAACTCGATATAGGATGGG 2435
 ||||| || | || | | | || | | | | | | | | | | | | | |
 Db 2344 ACGTTGCTCATGCATTT-----TATATCTTTAAATAAACAACCCAGTATGGAATGGG 2395

Qy 2436 AGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGA 2495
 | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 2396 AACCAATTATATATGAATTGAGTAGCTAGGCTATGCAGAAATTTCTGGAATCCTGAGAGG 2455

Qy 2496 ACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCAGGGGCCCCACACTCCGTGGTGA 2555
 | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 2456 ATAGTGGTTTATAGCAAAGTGTCTTAACCTTTCTCTTCTACCATTTCTCACAC----TGTTAA 2511

Qy 2556 GCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATG-CCATCCCTTCTTTT 2614
 |||| | | | | | | | | | | | | | | | | | | | | | | |
 Db 2512 GCCACTCCCAATACAAAGGGCGACCTAAACAAACTAGCAAAATGTTTTTCGCTTATCTC 2571

Qy 2615 TGTGTGGGGTCATGGGCTCCAAAAGCCAACGT 2646
 || |||| | | | | | | | | | | | | | | | | | | | | |
 Db 2572 TCGTGGATTTCATGGACTCCAACCCCAAAGT 2603

RESULT 7

AY196216

LOCUS AY196216 2284 bp mRNA linear ROD 01-JUN-2003

DEFINITION Mus musculus strain PERA/Ei ATP-binding cassette sub-family G member 8 (Abcg8) mRNA, complete cds.

ACCESSION AY196216

VERSION AY196216.1 GI:31322261

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2284)

AUTHORS Wittenburg,H., Lyons,M.A., Li,R., Churchill,G.A., Carey,M.C. and Paigen,B.

TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred Mice

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2284)

AUTHORS Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.

TITLE Direct Submission

JOURNAL Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

FEATURES Location/Qualifiers

source

1. .2284

/organism="Mus musculus"

/mol_type="mRNA"

/strain="PERA/Ei"

/db_xref="taxon:10090"

/chromosome="17"

/map="55 cm"

/sex="male"

/tissue_type="liver"

gene

1. .2284

/gene="Abcg8"

CDS

102. .2120

/gene="Abcg8"

/note="ATP-dependent canalicular cholesterol transporter; white subfamily"

/codon_start=1

/product="ATP-binding cassette sub-family G member 8"

/protein_id="AAO45096.1"

/db_xref="GI:31322262"

/translation="MAEKTKEETQLWNGTVLQDASGLQDSLFSSSEDNSLYFTYSGQS
NTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSQGMLA
IIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPN
LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKER
EVATVEKAQSLAALFLEKVQGFDDFLWKAELNTSTHTVSLTLTQDTCGTAVELP
GMIEQFSTLIRRIQISNDFRDLPTLLIHGSEACLSLIIGFLYGHGAKQLSFMDTAAL
LFMIGALIPFNVILDVVSCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVI
IYAMPIYWLTLNLRPVPELFLHFLLVLVVFCCRTMALAASAMLPFTFHMSSFFCNALY
NSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSIL
GDTMISAMDLNSHPLYAIYLIVIGISYGFLLFYLLSLKLIKQKSIQDW"

ORIGIN

Query Match 54.5%; Score 1454.2; DB 10; Length 2284;
 Best Local Similarity 79.8%; Pred. No. 1.3e-311;
 Matches 1760; Conservative 0; Mismatches 423; Indels 22; Gaps 3;

Qy	99	CATGGCCGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATAC	158
Db	101	CATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGC	160
Qy	159	CTCGGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTA	218
Db	161	TTCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTA	220
Qy	219	CAGTGGCCAGCCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTC	278
Db	221	CAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTC	280
Qy	279	TCAGGTCCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTG	338
Db	281	TCAGGTGCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAG	340
Qy	339	CCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGAT	398
Db	341	CCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGAT	400
Qy	399	GCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGG	458
Db	401	GCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGG	460
Qy	459	CCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTC	518
Db	461	CAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTAC	520
Qy	519	GCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAA	578
Db	521	GCCTCAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCAA	580
Qy	579	CTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCAGATGCGGCTGCCAGAACCTTCTC	638
Db	581	CCTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCAGATGCGCCTGCCAGGACCTTCTC	640
Qy	639	CCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTG	698
Db	641	CCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTG	700
Qy	699	CGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGTCGGGGGGTGAGCGCAGGAG	758
Db	701	CGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACG	760
Qy	759	AGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCAC	818
Db	761	AGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCAC	820
Qy	819	CTCTGGGCTCGACAGCTTCACAGCCCAACAACCTGGTGAAGACCTTGTCAGGCTGGCCAA	878
Db	821	TTCTGGCCTCGACAGCTTCACAGCCCAACAATCTGGTGACAACCTTGTCCTCCGCTGGCCAA	880

Qy	879	AGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTT	938
Db	881	GGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATT	940
Qy	939	TGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACAT	998
Db	941	TGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAAT	1000
Qy	999	GGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTT	1058
Db	1001	GGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTT	1060
Qy	1059	CTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGA	1118
Db	1061	CTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCCTGGA	1120
Qy	1119	GAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCT	1178
Db	1121	GAAGGCACAGTCTCTTGAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCT	1180
Qy	1179	ATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCC	1238
Db	1181	GTGGAAAGCTGAGGCAAAGGAAGTCAACACAAGCACCCACACAGTCAGCCTGACCCTCAC	1240
Qy	1239	ACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTAC	1298
Db	1241	ACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCAGGATGATAGAGCAGTTTTC	1297
Qy	1299	GACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCA	1358
Db	1298	CACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCA	1357
Qy	1359	TGGGGCGGAGGCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAG	1418
Db	1358	TGGGTGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGC	1417
Qy	1419	CATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTCATGATCGGTGCTCTCATCCC	1478
Db	1418	CAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCC	1477
Qy	1479	TTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTA	1538
Db	1478	TTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTA	1537
Qy	1539	TGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGA	1598
Db	1538	TGAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGA	1597
Qy	1599	GCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAA	1658
Db	1598	ATTGCCGGAGCACTGTGCCTACGTATCATCTACGCGATGCCATCTACTGGCTGACAAA	1657
Qy	1659	CCTGAGGCCAGGCCTCCAGCCCTTCTGCTGCACTTCTGCTGGTGTGGCTGGTGGTCTT	1718
Db	1658	CCTGCGGCCCCGTGCCTGAGCTCTTCTTCTACACTTCTGCTCGTGTGGTGGTGGTCTT	1717

Qy	1719	CTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCCTC	1778
Db	1718	CTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCTCTC	1777
Qy	1779	CTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTT	1838
Db	1778	CTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTT	1837
Qy	1839	GAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCCTTCCTGCGGTGGTGTTT	1898
Db	1838	GGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTT	1897
Qy	1899	TGAAGGGCTGATGAAGATTCACTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCT	1958
Db	1898	CTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTT	1957
Qy	1959	CACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCT	2018
Db	1958	CACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCCACT	2017
Qy	2019	CTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCGTGACTACGT	2078
Db	2018	CTATGCGATCTACCTCATGTGCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCT	2077
Qy	2079	GTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTG	2138
Db	2078	ATCCTTGAAGCTCATCAAACAGAAAGTCAATTCAAGACTGGTGATACTCAGCCTTGCTCTC	2137
Qy	2139	CCCGCTGGTGGGGGACCTGAGCAGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTC	2198
Db	2138	ACTGGCGG-----GACCCTTTTCCCGGGGCTGGCCACCCAGGAGGAGCCGGA	2185
Qy	2199	CTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACATCCGGCCAGGGTGCTG	2258
Db	2186	CTGGGGACAAGGCTCACACAGATCTCTCAG-----GCAGCAGCCACCTCTTAGTGCTG	2238
Qy	2259	CAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAGT	2303
Db	2239	CAGTGGCACAGGTCAGCCACAGGATGGCAGTAGAATAAAGACAGT	2283

TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone

Qy	279	TCAGGTCCCTTGGTTTGGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTG	338
Db	281	TCAGGTGCCTTGGTTTGGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAG	340
Qy	339	CCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGAT	398
Db	341	CCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGAT	400
Qy	399	GCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGG	458
Db	401	GCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGG	460
Qy	459	CCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTC	518
Db	461	CAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAACGGGCAACCCAGTAC	520
Qy	519	GCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAA	578
Db	521	GCCTCAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCAA	580
Qy	579	CTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTC	638
Db	581	CCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTC	640
Qy	639	CCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTG	698
Db	641	CCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTG	700
Qy	699	CGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAG	758
Db	701	CGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGTGAGCGCCGACG	760
Qy	759	AGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCAC	818
Db	761	AGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCAC	820
Qy	819	CTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAA	878
Db	821	TTCTGGCCTCGACAGCTTCACAGCCCACAACCTGGTGACAACCTTGTCCCGCCTGGCCAA	880
Qy	879	AGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTT	938
Db	881	GGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATT	940
Qy	939	TGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACAT	998
Db	941	TGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAAT	1000
Qy	999	GGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTT	1058
Db	1001	GGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCAGACTT	1060
Qy	1059	CTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGA	1118
Db	1061	CTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGA	1120

Qy	1119	GAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCT	1178
Db	1121	GAAGGCACAGTCTCTTGCAGCCCTGTTCTTAGAAAAAGTGCAAGGCTTTGATGACTTTCT	1180
Qy	1179	ATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCC	1238
Db	1181	GTGGAAAGCTGAGGCAAAGGAACCTCAACACAAGCACCCACACAGTCAGCCTGACCCTCAC	1240
Qy	1239	ACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCGTGGGGCGGTGCAGCAGTTTAC	1298
Db	1241	ACAGGACACTGACTG---TGGGACTGCTGCTGAGCTGCCCGGGATGATAGAGCAGTTTTC	1297
Qy	1299	GACGCTGATCCGTCGTGATTTCCAACGACTTCCGAGACCTGCCACCCCTCCTCATCCA	1358
Db	1298	CACCCTGATCCGTCGTGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCA	1357
Qy	1359	TGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAG	1418
Db	1358	TGGGTCGGAAGCCTGCCTGATGTCCCTCATCATCGGCTTCCTTTACTACGGCCATGGGGC	1417
Qy	1419	CATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCC	1478
Db	1418	CAAGCAGCTCTCCTTCATGGACACGGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCC	1477
Qy	1479	TTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTA	1538
Db	1478	TTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCCTCGGAGAGGTCAATGCTGTACTA	1537
Qy	1539	TGAACTGGAAGACGGGCTGTACACCCTGGTCCATATTTCTTTGCCAAGATCCTCGGGGA	1598
Db	1538	TGAGCTGGAAGACGGGCTGTACACTGCTGGTCCCTATTTCTTTGCCAAGATCCTAGGAGA	1597
Qy	1599	GCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAA	1658
Db	1598	ATTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAA	1657
Qy	1659	CCTGAGGCCAGGCCTCCAGCCCTTCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTT	1718
Db	1658	CCTGCGGCCCCTGCCTGAGCTCTTCCTTCTACACTTACTGCTTGTGTGGTTGGTGGTCTT	1717
Qy	1719	CTGTTGCAGGATTATGGCCCTGGCCGCCCGGCCCTGCTCCCCACCTTCCACATGGCCTC	1778
Db	1718	CTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTC	1777
Qy	1779	CTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTT	1838
Db	1778	CTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACCGCCGGCTTCATGATAAACTT	1837
Qy	1839	GAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCCTTCCTGCGGTGGTGT	1898
Db	1838	GGACAACCTGTGGATAGTGCCTGCATGGATATCCAAGCTGTCGTTCTCCGGTGGTGT	1897
Qy	1899	TGAAGGGCTGATGAAGATTCAATTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCT	1958
Db	1898	CTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTT	1957
Qy	1959	CACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCT	2018

```

      ||||| || | || ||||| | || ||||| ||||| || ||||| || || ||
Db      1958 CACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACT 2017
Qy      2019 CTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGT 2078
      ||| || ||||| ||||| ||||| ||| ||||| ||||| || ||||| ||| |
Db      2018 CTATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCT 2077
Qy      2079 GTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTG 2138
      ||||| | | ||||| ||||| ||| ||||| ||||| || | | |||
Db      2078 ATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGATACTCAGCCTTGCTCTC 2137
Qy      2139 CCCGCTGGTGGGGGACCTGAGCAGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCTTC 2198
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      2138 ACTGGCG-----GGACCCTTTCCCGGGGCTGGCCACCCCAGGAGGCCGGA 2184
Qy      2199 CTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTG 2258
      ||||| || | ||| ||| ||| || || || | | |||||
Db      2185 CTGGGGACAAGGCTCACACAGATCTCTCAG-----GCAGCAGCCACCTCTTAGTGCTG 2237
Qy      2259 CAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAA 2297
      ||||| ||||| ||||| ||||| |||||
Db      2238 CAGTGGCACAGGTCAGCCACAGGATGGCAGTAAATAAA 2276

```

RESULT 9

AX685731

LOCUS AX685731 2019 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 3 from Patent WO02081691.

ACCESSION AX685731

VERSION AX685731.1 GI:29371740

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 3 17-OCT-2002; Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM (US)

FEATURES

source

Location/Qualifiers

1. .2019

/organism="Mus musculus"

/mol_type="unassigned DNA"

/db_xref="taxon:10090"

CDS

1. .2019

/note="unnamed protein product; mouse ABCG8 (mABCG8)"

/codon_start=1

/protein_id="CAD86571.1"

/db_xref="GI:29371741"

/db_xref="REMTREMBL:CAD86571"

/translation="MAEKTKEETQLWNGTVLQDASGLQDSLFSSSEDNSLYFTYSGQSN
NTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLA
IIGSSGCGRASLLDVITGRGHGGMKSGQIWIWGPSTPQLVRKCVAHVRQHDQLLPN
LTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDI

FRLFDLVLLMTSGTFIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKER
EVATVEKAQSLAALFLEKVQGGDFDLWKAEAKELNTSTHTVSLTLTQDTCGTAVELP
GMIEQFSTLIRROI SNDFRDLP TLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAAL
LFMIGALIPFNVIDLVVSKCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVI
IYAMPIYWLTLNLRPVPELFLHLHLLVWL VVFCRTMALAASAMLP TFHMSSFFCNALY
NSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSIL
GDTMISAMDLSHPLYAIYLIVIGISYGFLFLYYLSLKLKQKSIQDW"

Query Match 53.6%; Score 1430; DB 6; Length 2019;
Best Local Similarity 82.0%; Pred. No. 3.1e-306;
Matches 1659; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

Qy	760	GTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC	819
Db	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Qy	820	TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCAGGCTGGCCAAA	879
Db	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCCCTGGCCAAG	780
Qy	880	GGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTT	939
Db	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Qy	940	GATCTGGTCCTCCTGATGACGTCTGGCACCCCCATCTACTTAGGGGCGGCCAGCACATG	999
Db	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Qy	1000	GTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTC	1059
Db	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTC	960
Qy	1060	TATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAG	1119
Db	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCCTGGAG	1020
Qy	1120	AAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCCTGACTTAGATGACTTTCTA	1179
Db	1021	AAGGCACAGTCTCTTGACAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Qy	1180	TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCA	1239
Db	1081	TGGAAAGCTGAGGCAAAGGAACCTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Qy	1240	CTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG	1299
Db	1141	CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTC	1197
Qy	1300	ACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCAT	1359
Db	1198	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCAT	1257
Qy	1360	GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGC	1419
Db	1258	GGGTGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1317
Qy	1420	ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCT	1479
Db	1318	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1377
Qy	1480	TTCAACGTCAATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT	1539
Db	1378	TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT	1437
Qy	1540	GAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG	1599
Db	1438	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA	1497
Qy	1600	CTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAAC	1659

Db	1498	TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1557
Qy	1660	CTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC	1719
Db	1558	CTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC	1617
Qy	1720	TGTTGCAGGATTATGGCCCTGGCCGCCCGGCCCTGCTCCCCACCTTCCACATGGCCTCC	1779
Db	1618	TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCCTCC	1677
Qy	1780	TTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG	1839
Db	1678	TTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG	1737
Qy	1840	AGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTTC	1899
Db	1738	GACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTC	1797
Qy	1900	GAAGGGCTGATGAAGATTCAAGTTCAGTCAGAGAAGAACTTATAAAATGCCTCTCGGGAACCTC	1959
Db	1798	TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC	1857
Qy	1960	ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTC	2019
Db	1858	ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC	1917
Qy	2020	TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCCTGTACTACGTG	2079
Db	1918	TATGCGATCTACCTCATTGTATCGGCATCAGCTACGGCTTCCTGTTCCCTGTACTATCTA	1977
Qy	2080	TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA	2121
Db	1978	TCCTTGAAGCTCATCAAACAGAAAGTCAATTCAAGACTGGTGA	2019

RESULT 10

AC084265

LOCUS AC084265 127066 bp DNA linear PRI 11-DEC-2001

DEFINITION Homo sapiens chromosome 2, clone CTB-2367F13, complete sequence.

ACCESSION AC084265

VERSION AC084265.4 GI:17488659

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 2, clone CTB-2367F13

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,

FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
 Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (18-OCT-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (24-AUG-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 4 (bases 1 to 127066)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,

Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Dec 11, 2001 this sequence version replaced gi:15284200.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11578
 Center clone name: 2367_F_13
 -----.

FEATURES	Location/Qualifiers
source	1. .127066 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="2" /map="2" /clone="CTB-2367F13" /clone_lib="CITB Human BAC"
repeat_region	complement(8. .170) /rpt_family="MER47A"
repeat_region	171. .468 /rpt_family="AluSx"
repeat_region	complement(469. .516) /rpt_family="MER47A"
repeat_region	988. .1049 /rpt_family="MIR"
repeat_region	complement(1294. .1448) /rpt_family="L1ME4A"
repeat_region	complement(2662. .2954) /rpt_family="AluSx"
repeat_region	4049. .4431 /rpt_family="L2"
unsure	5261. .5269 /note="<30 qual SNGL region"
unsure	7192. .7202 /note="<30 qual SNGL region"
repeat_region	7310. .7472 /rpt_family="MIR"
repeat_region	7488. .7582 /rpt_family="MIR"

repeat_region	7589. .7628
	/rpt_family="(TTG)n"
repeat_region	complement(7631. .7781)
	/rpt_family="AluSg/x"
repeat_region	7791. .7922
	/rpt_family="MIR"
repeat_region	complement(7977. .8300)
	/rpt_family="AluJb"
repeat_region	9044. .9343
	/rpt_family="AluSq"
repeat_region	10315. .10344
	/rpt_family="AT_rich"
repeat_region	10355. .10681
	/rpt_family="AluJo"
repeat_region	10683. .10993
	/rpt_family="AluSx"
repeat_region	complement(12221. .12282)
	/rpt_family="MIR3"
repeat_region	complement(12306. .12449)
	/rpt_family="MIR"
repeat_region	complement(13008. .13189)
	/rpt_family="MER33"
repeat_region	complement(13190. .13471)
	/rpt_family="AluJo"
repeat_region	complement(13472. .13612)
	/rpt_family="MER33"
repeat_region	13899. .13922
	/rpt_family="GC_rich"
repeat_region	complement(14184. .14250)
	/rpt_family="L2"
repeat_region	14552. .14630
	/rpt_family="MER5A"
repeat_region	complement(14809. .15100)
	/rpt_family="AluSx"
repeat_region	complement(15363. .15679)
	/rpt_family="AluY"
repeat_region	complement(15681. .15979)
	/rpt_family="AluSx"
repeat_region	16292. .16388
	/rpt_family="L2"
repeat_region	16392. .16508
	/rpt_family="MLT1F"
repeat_region	complement(16538. .16616)
	/rpt_family="LTR37B"
repeat_region	16618. .16687
	/rpt_family="Alu"
repeat_region	complement(16988. .17104)
	/rpt_family="L2"
repeat_region	17540. .17895
	/rpt_family="MLT1A1"
repeat_region	complement(17911. .18209)
	/rpt_family="AluSq"
repeat_region	18487. .18680
	/rpt_family="LTR16A1"
repeat_region	18802. .19026
	/rpt_family="AluJo"
repeat_region	complement(19092. .19390)

Db 60098 CCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATAT 60157
 Qy 2414 AGGCAACTCGATATAGGATGGGAGCAAAGTAGGAATGAATTGGGTAGCTAGACTGTGCAG 2473
 |||||
 Db 60158 AGGCAACTCGATATAGGATGGGAGCAAAGTAGGAATGAATTGGGTAGCTAGACTGTGCAG 60217
 Qy 2474 GAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCA 2533
 |||||
 Db 60218 GAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCA 60277
 Qy 2534 GGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAG 2593
 |||||
 Db 60278 GGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAG 60337
 Qy 2594 CAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAAT 2653
 |||||
 Db 60338 CAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAAT 60397
 Qy 2654 TAAAAAATGTATTGAGC 2669
 |||||
 Db 60398 TAAAAAATGTATTGAGC 60413

RESULT 11

AC087053

LOCUS AC087053 182261 bp DNA linear HTG 27-MAR-2003

DEFINITION Homo sapiens chromosome 2 clone RP11-959M3 map 2, 9 unordered pieces.

ACCESSION AC087053

VERSION AC087053.13 GI:25140148

KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLLTOP; HTGS_CANCELLED.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 182261)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 2, clone RP11-959M3

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 182261)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
 Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
 Lehoczký,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
 Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
 Rieback,M., Riley,R., Rogov,P., Roman,J., Rosetti,M., Roy,A.,
 Santos,R., Schauer,S., Seaman,S., Severy,P., Sougnez,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N.,

Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M.,
Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (30-NOV-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 182261)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faró,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Nov 20, 2002 this sequence version replaced gi:25046635.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11589
Center clone name: 959_M_3

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 78629: contig of 78629 bp in length
* 78630 78729: gap of 100 bp
* 78730 94580: contig of 15851 bp in length
* 94581 94680: gap of 100 bp
* 94681 112894: contig of 18214 bp in length
* 112895 112994: gap of 100 bp
* 112995 140439: contig of 27445 bp in length

[illegible]

Qy 2234 CAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAA 2293
 |||
 Db 7350 CAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAA 7409

Qy 2294 TAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAA 2353
 |||
 Db 7410 TAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAA 7469

Qy 2354 CCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATAT 2413
 |||
 Db 7470 CCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATAT 7529

Qy 2414 AGGCAACTCGATATAGGATGGGAGCAAAGTGAATTTGGGTAGCTAGACTGTGCAG 2473
 |||
 Db 7530 AGGCAACTCGATATAGGATGGGAGCAAAGTGAATTTGGGTAGCTAGACTGTGCAG 7589

Qy 2474 GAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCA 2533
 |||
 Db 7590 GAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCA 7649

Qy 2534 GGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAG 2593
 |||
 Db 7650 GGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAG 7709

Qy 2594 CAAGATGCCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAAT 2653
 |||
 Db 7710 CAAGATGCCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAAT 7769

Qy 2654 TAAAAATGTATTGAGC 2669
 |||
 Db 7770 TAAAAATGTATTGAGC 7785

RESULT 12

AC108476

LOCUS AC108476 139342 bp DNA linear PRI 16-APR-2002

DEFINITION Homo sapiens BAC clone RP11-1413K20 from 2, complete sequence.

ACCESSION AC108476

VERSION AC108476.5 GI:19807988

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 139342)

AUTHORS Sulston, J.E. and Waterston, R.

TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792

PUBMED 9847074

REFERENCE 2 (bases 1 to 139342)

AUTHORS Harkins, C., Haakenson, W. and Doebber, A.

TITLE The sequence of Homo sapiens BAC clone RP11-1413K20

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 139342)

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (27-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 4 (bases 1 to 139342)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (20-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 5 (bases 1 to 139342)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 6 (bases 1 to 139342)

AUTHORS Waterston,R.

TITLE Direct Submission

JOURNAL Submitted (16-APR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT On Mar 29, 2002 this sequence version replaced gi:18767626.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH1413K20
-----.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality ≥ 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-489K22, 2000 bp overlap.
Actual end is at base position 139342 of RP11-1413K20.

The region between 132012 and 132017 is covered only by a pcr
product of clone DNA.

FEATURES	Location/Qualifiers
source	1. .139342 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="2" /map="2" /clone="RP11-1413K20" /clone_lib="RPCI-11"
misc_feature	55. .655 /note="match to EST AA203458 (NID:g1799169) zx58b04.r1"
misc_feature	93. .286 /note="match to EST AV689089 (NID:g10290952)"
misc_feature	93. .286 /note="similar to Mus musculus EST AI597378 (NID:g4606426) vj29c06.y1"
misc_feature	93. .279 /note="match to EST AV660973 (NID:g9881987)"
misc_feature	318. .653 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
misc_feature	372. .633 /note="similar to Homo sapiens EST T97887 (NID:g747232) ye58h05.r1"
misc_feature	706. .708 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
misc_feature	706. .707 /note="similar to Homo sapiens EST T97887 (NID:g747232) ye58h05.r1"
repeat_region	847. .1139 /rpt_family="Alu"
misc_feature	1867. .2047 /note="match to EST T39945 (NID:g647612) ya13g04.r1"
repeat_region	2234. .2616 /rpt_family="L2"
misc_feature	2983. .3121 /note="match to EST AV689089 (NID:g10290952)"
misc_feature	2983. .3121 /note="similar to Mus musculus EST AI597378 (NID:g4606426) vj29c06.y1"
misc_feature	3044. .3121 /note="match to EST T86384 (NID:g714736) yd77b08.r1"
misc_feature	4099. .4304 /note="match to EST T86384 (NID:g714736) yd77b08.r1"
misc_feature	4099. .4283 /note="match to EST AV689089 (NID:g10290952)"
misc_feature	4401. .4618 /note="similar to Mus musculus EST BF162656 (NID:g11042879)"
misc_feature	4405. .4454

```

misc_feature      /note="match to EST T86384 (NID:g714736) yd77b08.r1"
                  4724. .5110
                  /note="similar to Homo sapiens EST AV656623
                  (NID:g9877637)"
misc_feature      5075. .5204
                  /note="similar to Mus musculus EST BF162656
                  (NID:g11042879)"
repeat_region     5495. .5657
                  /rpt_family="MIR"
repeat_region     5673. .5767
                  /rpt_family="MIR"
repeat_region     5774. .5813
                  /rpt_family="(TTG)n"
repeat_region     5816. .5958
                  /rpt_family="Alu"
repeat_region     5976. .6091
                  /rpt_family="MIR"
repeat_region     6162. .6485
                  /rpt_family="Alu"
misc_feature      6351. .6373
                  /note="match to EST AA228345 (NID:g1849916) nc39d04.sl"
misc_feature      6352. .6364
                  /note="match to EST AI431309 (NID:g4302284) ar55b01.x1"
misc_feature      6352. .6364
                  /note="match to EST AI469772 (NID:g4331862) tm20f11.x1"
misc_feature      6353. .6367
                  /note="match to EST AI241685 (NID:g3837082) qu70f06.x1"
misc_feature      6568. .6707
                  /note="similar to Mus musculus EST BF162656
                  (NID:g11042879)"
misc_feature      6649. .6707
                  /note="similar to Mus musculus EST BB598373
                  (NID:g16450340)"
repeat_region     7229. .7528
                  /rpt_family="Alu"
misc_feature      7940. .8549
                  /note="similar to EST BM725726 (NID:g19047059)"
misc_feature      8169. .8305
                  /note="similar to Mus musculus EST BF162656
                  (NID:g11042879)"
misc_feature      8169. .8301
                  /note="similar to Mus musculus EST BB598373
                  (NID:g16450340)"
repeat_region     8500. .8529
                  /rpt_family="AT_rich"
repeat_region     8540. .8868
                  /rpt_family="Alu"
repeat_region     8870. .9180
                  /rpt_family="Alu"
repeat_region     10493. .10636
                  /rpt_family="MIR"
repeat_region     11195. .11376
                  /rpt_family="MER1_type"
repeat_region     11377. .11658
                  /rpt_family="Alu"
repeat_region     11659. .11799
                  /rpt_family="MER1_type"

```

```
misc_feature      11955..12053  
                  /note="similar to Mus musculus EST BB598373  
                  (NID:g16450340)"  
misc_feature      11994..12053  
                  /note="similar to Mus musculus EST AA239884 (NID:g1863923)  
                  mx8ld01.r1"  
repeat_region     12086..12109
```

Query Match 27.1%; Score 722.4; DB 9; Length 139342;
Best Local Similarity 89.8%; Pred. No. 2e-149;
Matches 823; Conservative 0; Mismatches 6; Indels 87; Gaps 1;

Qy	1841	GCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCCTGCGGTGGTGTTTTG	1900
Db	57732	GCTGTCTGTCTCCAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCCTGCGGTGGTGTTTTG	57791
Qy	1901	AAGGGCTGATGAAGATTCAAGTTCAGCAGAGAAGAACTTATAAAATGCCTCTCGGGAACCTCA	1960
Db	57792	AAGGGCTGATGAAGATTCAAGTTCAGCAGAGAAGAACTTATAAAATGCCTCTCGGGAACCTCA	57851
Qy	1961	CCATCGCGGTCTCAGGAGATAAA-----	1983
Db	57852	CCATCGCGGTCTCAGGAGATAAAGTAAGCGGGGAAGGCCTCGGGTTCTAAATTATTGGAC	57911
Qy	1984	-----ATCCTCAGTG	1993
Db	57912	GTCCGGCTTTCCATCCTCCTCATGAGCCCACTGCATGTCTGTGTCTCCAGATCCTCAGTG	57971
Qy	1994	CCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCG	2053
Db	57972	TCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCG	58031
Qy	2054	GTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAG	2113
Db	58032	GTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAG	58091
Qy	2114	ACTGGTGATTACAGCCAGACGTCTGCCCGCTGGTGGGGACCTGAGCAGACCCTTCAACT	2173
Db	58092	ACTGGTGATTACAGCCAGACGTCTGCCCGCTGGTGGGGACCTGAGCAGACCCTTCAACT	58151
Qy	2174	GCACTCCCTCCTCAGGAGCCCTTCTGTTGGGACAGTGAGGACAATGACCCTACAGATGCT	2233
Db	58152	GCACTCCCTCCTCAGGAGCCCTTCTGTTGGGACAGTGAGGACAATGACCCTACAGATGCT	58211
Qy	2234	CAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAA	2293
Db	58212	CAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAA	58271
Qy	2294	TAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAA	2353
Db	58272	TAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAA	58331
Qy	2354	CCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATAT	2413
Db	58332	CCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATAT	58391
Qy	2414	AGGCAACTCGATATAGGATGGGAGCAAAGTGAATGAGGATGAGTGTGCAG	2473

```

|||||
Db      58392 AGGCAACTCGATATAGGATGGGAGCAAAGTAGGAATGAATTGGGTAGCTAGACTGTGCAG 58451
Qy      2474 GAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCA 2533
|||||
Db      58452 GAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCA 58511
Qy      2534 GGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAG 2593
|||||
Db      58512 GGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAG 58571
Qy      2594 CAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAAT 2653
|||||
Db      58572 CAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAAT 58631
Qy      2654 TAAAAATGTATTGAGC 2669
|||||
Db      58632 TAAAAATGTATTGAGC 58647

```

RESULT 13
F351812S13

LOCUS F351812S13 2201 bp DNA linear PRI 10-AUG-2001

DEFINITION Homo sapiens sterolin-2 (ABCG8) gene, exon 13 and complete cds.

ACCESSION AF351824

VERSION AF351824.1 GI:15146443

KEYWORDS .

SEGMENT 13 of 13

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2201)

AUTHORS Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H.,
Ose, L., Stalenhoeft, A.F., Mietinnen, T., Bjorkhem, I., Bruckert, E.,
Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Srivastava, A. and
Patel, S.B.

TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterolin-1 and
sterolin-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)

MEDLINE 21344600

PUBMED 11452359

REFERENCE 2 (bases 1 to 2201)

AUTHORS Lu, K.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers

source 1..2201

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/map="between D2S2294 and D2S2298"

/clone="1081G2; 328I4"

Qy 2102 AACCAAGTCAAGACTGGTGATTACGCCAGACGCTGCCCCGCTGGTGGGGGACCTGAGCA 2161
 |||
 Db 163 AACCAAGTCAAGACTGGTGATTACGCCAGACGCTGCCCCGCTGGTGGGGGACCTGAGCA 222
 Qy 2162 GACCCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGAC 2221
 |||
 Db 223 GACCCCTTCAACTGCACTCCCTCCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGAC 282
 Qy 2222 CCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGG 2281
 |||:|||
 Db 283 CCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCRGTGGCACAGACCAGCCACAGG 342
 Qy 2282 ATGGCAGTAGAATAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATG 2341
 |||
 Db 343 ATGGCAGTAGAATAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATG 402
 Qy 2342 ACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGAT 2401
 |||
 Db 403 ACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGAT 462
 Qy 2402 ATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGC 2461
 |||
 Db 463 ATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGC 522
 Qy 2462 TAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGG 2521
 |||:|||
 Db 523 TAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTASCTAGCAGATTTGG 582
 Qy 2522 CTTTCATCTTCCAGGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCT 2581
 |||
 Db 583 CTTTCATCTTCCAGGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCT 642
 Qy 2582 -AAGATGTACCAGCAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGC 2640
 |||
 Db 643 NAAGATGTACCAGCAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGC 702
 Qy 2641 CAACGTGAACAA-TTAAAAATGTATTGAGC 2669
 |||
 Db 703 CAACGTGAACAANTTAAAAATGTATTGAGC 732

RESULT 14

AC146466

LOCUS AC146466 185045 bp DNA linear HTG 15-AUG-2003
 DEFINITION Callithrix jacchus clone CH259-274K20, WORKING DRAFT SEQUENCE, 3
 ordered pieces.

ACCESSION AC146466

VERSION AC146466.1 GI:33667132

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE Callithrix jacchus (white-tufted-ear marmoset)

ORGANISM Callithrix jacchus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
 Callithrix.

REFERENCE 1 (bases 1 to 185045)

AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,

TITLE Peng,Z., Malinov,I. and Rubin,E.M.
 JOURNAL Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 185045)
 AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
 Peng,Z., Malinov,I. and Rubin,E.M.
 TITLE Direct Submission
 JOURNAL Submitted (15-AUG-2003) Genome Sciences, Lawrence Berkeley National
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

COMMENT

Sequence Produced by Berkeley PGA
 Web site: <http://pga.lbl.gov>
 Center Code: PGABERK
 Center Project Name: J027
 Bac Clone Name: CH259-274K20

This sequence has been compared to sequences of other species
 using Vista (<http://www-gsd.lbl.gov/VISTA>). The results can be
 viewed at:
http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=ABCG5

The order-orientation of the draft sequence was accomplished by
 using:
 Avid (<http://baboon.math.berkeley.edu/mavid>),
 Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

Summary Statistics:

Sequencing vector: Plasmid; pUC18
 Chemistry: Dye-terminator Big Dye
 Assembly program: Phrap version 0.990329.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 49109: contig of 49109 bp in length
 * 49110 49209: gap of unknown length
 * 49210 57420: contig of 8211 bp in length
 * 57421 57520: gap of unknown length
 * 57521 185045: contig of 127525 bp in length.

FEATURES

source

Location/Qualifiers
 1. .185045
 /organism="Callithrix jacchus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9483"
 /clone="CH259-274K20"

ORIGIN

Query Match 21.8%; Score 581.2; DB 2; Length 185045;
 Best Local Similarity 83.0%; Pred. No. 4.1e-118;
 Matches 762; Conservative 0; Mismatches 63; Indels 93; Gaps 5;

Qy 1841 GCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGCCTTCCTGCGGTGGTGTGTTTG 1900
 || ||| |||||||||||||||||||||||||||||||||||
 Db 139563 GCTCTCTGTCTCCAGTGCCCGCGTGGATTTCCAAAGTGCCTTCCTGCGATGGTGTGTTTG
 139622

Qy 1901 AAGGGCTGATGAAGATTCAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCA 1960
 ||||||||||||||||||||||||| ||| ||||||||||||||| || ||| |||
 Db 139623 AAGGGCTGATGAAGATTCAGTTCAGCAGCAGAGCTTATAAAATGCCTCTGGGCAACTTCA
 139682

Qy 1961 CCATCGCGGTCTCAGGAGATAAA----- 1983
 ||||| | ||| |||||||||
 Db 139683 CCATCCCAGTCCCAGGAGATAAAGTAAGCGGGGAAGGCCTCAGGTTCTAAATGACTGGAT
 139742

Qy 1984 -----ATCCTCAGTG 1993
 |||||||||
 Db 139743 GTCCGGCTGCCTGTCCTCCTAATGAGCCCACTGCATGTCTGTGTCTCCAGATCCTCAGTG
 139802

Qy 1994 CCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCG 2053
 ||||||||| ||||||||||||||| |||||||||||||||
 Db 139803 CCATGGAGCTGAACTCGTACCCTCTCTACGTCATCTACCTCATCGTCATTGGCCTCAGCG
 139862

Qy 2054 GTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAG 2113
 ||||||||||||||||||||||| |||||||||||||||
 Db 139863 GTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAG
 139922

Qy 2114 ACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGCAGACCCTTCAACT 2173
 ||||||||| ||| | ||||||||||||||| ||| |||||||||||
 Db 139923 ACTGGTGATTTCATGCCGGGTGCCTGCCCGCTGGTGGGGCACCCAAGCAGACCCTTCAACT
 139982

Qy 2174 GCACTCCCTCCTCAGGAGCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGATGCT 2233
 ||||| ||||| ||||||||||||||| ||||||||||||||| ||| |||
 Db 139983 GCACTCCTTCCTCAGAAGCCCCTTCCTGGGGACAATGAGGACAATGACCCTA-AGAAGCT
 140041

Qy 2234 CAGCTACATCCGGCCCAGGGTGTGTCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAA 2293
 ||||||||||||| ||||||||||||||| |||||||||||||||
 Db 140042 CAGCTACATCCGGCCCACAGTGTGTCAGTGGCACAGACCAACCACAGGATGGCAGTAGAA
 140101

Qy 2294 TAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTG--CGATGACTGGGAGAA 2351
 ||||||||| ||||||||||||||| ||||||||||||| || |||
 Db 140102 TAAAGACAGTAGAAAGGGATTTCTGCTCACTGGCAGGAGACTGATGACTGGGAGTGAGAA
 140161

Qy 2352 AACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTAT 2411
 ||||||||| ||||| ||||||||||||||| ||||||||||||| |||||
 Db 140162 AACCTGCACTCAGTGGCGCCTACAACGTTGCTAATTTATTTCTTTTGATATGTGCTTAT
 140221

Qy 2412 ATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGC 2471

```

|||||
Db      140222 ATAGGCAACTCGATATAGGATGGGAGCAAAGTAGGAATAAATTGAGTAGCTAGAC--TGC
140279

Qy      2472 AGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTC 2531
|||||
Db      140280 AGGAATTGTTAGAACCCAGAGAGAATAATAAAAGTAGCTAGCAGATCTGGCCTCATCTTC
140339

Qy      2532 CAGGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACC 2591
|||||
Db      140340 CAGGGGCCCCACACTTAGTGG-GAGCTATCATCAATACAGAAAGTGATCTAAGATGTACC
140398

Qy      2592 AGCAAGATGCCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACA 2651
|||||
Db      140399 AGCAAGATGCCACCCCTTCTTTTCGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACA
140458

Qy      2652 ATTAAAAATGTATTGAGC 2669
|||||
Db      140459 ATTAAAAATTTATTGAGC 140476

```

RESULT 15

AC146464/c

LOCUS AC146464 202533 bp DNA linear HTG 19-AUG-2003

DEFINITION Saimiri sciureus clone CH254-84A11, WORKING DRAFT SEQUENCE.

ACCESSION AC146464

VERSION AC146464.1 GI:33636782

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE Saimiri sciureus (common squirrel monkey)

ORGANISM Saimiri sciureus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;
Saimiri.

REFERENCE 1 (bases 1 to 202533)

AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 202533)

AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.

TITLE Direct Submission

JOURNAL Submitted (14-AUG-2003) Genome Sciences, Lawrence Berkeley National
Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

REFERENCE 3 (bases 1 to 202533)

AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,
Peng, Z., Malinov, I. and Rubin, E.M.

TITLE Direct Submission

JOURNAL Submitted (19-AUG-2003) Genome Sciences, Lawrence Berkeley National
Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

COMMENT

Sequence Produced by Berkeley PGA
Web site: <http://pga.lbl.gov>
Center Code: PGABERK

Center Project Name: S030
Bac Clone Name: CH254-84A11

This sequence has been compared to sequences of other species using Vista (<http://www-gsd.lbl.gov/VISTA>). The results can be viewed at:
http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=ABCG5

The order-orientation of the draft sequence was accomplished by using:
Avid (<http://baboon.math.berkeley.edu/mavid>),
Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

Summary Statistics:

Sequencing vector: Plasmid; pUC18

Chemistry: Dye-terminator Big Dye

Assembly program: Phrap version 0.990329.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
 source
 1 202533: contig of 202533 bp in length.
 Location/Qualifiers
 1. .202533
 /organism="Saimiri sciureus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9521"
 /clone="CH254-84A11"

ORIGIN

Query Match 21.6%; Score 576.4; DB 2; Length 202533;
Best Local Similarity 82.7%; Pred. No. 4.7e-117;
Matches 759; Conservative 0; Mismatches 66; Indels 93; Gaps 5;

```
Qy      1841 GCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGCCTTCCTGCGGTGGTGTGTTTG 1900
          ||   |||   |||||
Db      10147 GCTCTCTGTCTCCAGTGCCCGCGTGGATTTCCAAAGTGCCTTCCTGCGGTGGTGTGTTTG 10088

Qy      1901 AAGGGCTGATGAAGATTCAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCA 1960
          |||||
Db      10087 AAGGGCTGATGAAGATTCAGTTCAGCAGCAGAGCTTATAAAATGCCTCTGGGCAACTTCA 10028

Qy      1961 CCATCGCGGTCTCAGGAGATAAA----- 1983
          |||||
Db      10027 CCATCCCGGTCCCAGGAGATAAAGTAAGCGGGGAAGGCCTCAGGTTCTAAATGACTGGAT 9968

Qy      1984 -----ATCCTCAGTG 1993
          |||||
Db      9967 GTCCGGCTGCCTGTCTCCTAACGAGCCCCTGCATGTCTGTGTCTCCAGATCCTCAGTG 9908
```

Qy	1994	CCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCG	2053
Db	9907	CCATGGAGCTGAACTCGTACCCTCTCTACGCCATCTACCTTATCGTCATTGGTCTCAGCG	9848
Qy	2054	GTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAG	2113
Db	9847	GTGGCTTCATGGTCCTGTACTACATGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAG	9788
Qy	2114	ACTGGTGATTACAGCCAGACGTCTGCCCCTGGTGGGGGACCTGAGCAGACCCTTCAACT	2173
Db	9787	ACTGGTGATTACATGCCGGGCGCCTGCCCACTGGTGGGGAACCCGAGCAGACCCTTCAACT	9728
Qy	2174	GCACTCCCTCCTCAGGAGCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGATGCT	2233
Db	9727	GCACTTTTTCCTCAGGAGCCCCTTCCTGGGGACAATGAGGACAATGACCCTA-AGAAGCT	9669
Qy	2234	CAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAA	2293
Db	9668	CAGCTACATCCGGCCCACAGTGCTGCAGTGGCACAGGCCAGCCACAGGATGGCAGTAGAA	9609
Qy	2294	TAAAGACAGTCGAAAGGGATTTCTGCTCACTGGCAGGAGACTG--CGATGACTGGGAGAA	2351
Db	9608	TAAAGACAGTCGAAAGGGATTTCTGCTCTCTGGCAGGAGACTGATGACTGGGAGTGAGAA	9549
Qy	2352	AACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTAT	2411
Db	9548	AACCTGCACTCAGTGGCGCCACAACGTTGCTAATTTATTTCTTTTGATATGTGCTTAT	9489
Qy	2412	ATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGC	2471
Db	9488	ATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATAAATTGAGTAGCTAGAC--TGC	9431
Qy	2472	AGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTC	2531
Db	9430	AGGAATTGTTAGAACCCAGAGAGAACAATAACAGTAGCTAGCAGATCTGATCTCATCTTC	9371
Qy	2532	CAGGGGCCCCACACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACC	2591
Db	9370	CAGGGGCCCCACACTTCGTGG-GAGCCATCATCAATACAGAACGTGACCTAAGATGTACC	9312
Qy	2592	AGCAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACA	2651
Db	9311	AGCAAGATGCCACCCCTTCTGTTCTGTGGGGTAATGGGCTCCAAAAGCCAACGTGAACA	9252
Qy	2652	ATTAAAAATGTATTGAGC	2669
Db	9251	ATTAAAAATTTATTGGGC	9234

Search completed: February 26, 2004, 06:21:30
 Job time : 6869.55 secs

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:39:18 ; Search time 675.985 Seconds
(without alignments)
16773.223 Million cell updates/sec

Title: US-09-989-981A-7
Perfect score: 2669
Sequence: 1 gtgtccctgctccaggaaac.....caattaaaaatgtattgagc 2669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2669	100.0	2669	7	AAD48883	Aad48883 Human ABC
2	1680.6	63.0	3239	6	ABK83218	Abk83218 Human tra
3	1499	56.2	2564	6	ABN90022	Abn90022 Mouse clo
4	1430	53.6	2019	7	AAD48881	Aad48881 Mouse ABC
5	291.6	10.9	580	4	AAH98911	Aah98911 Arabidops
6	203.6	7.6	1920	6	ABK51681	Abk51681 DNA encod
7	203.6	7.6	2340	6	AAD22009	Aad22009 Human sit

	8	203.6	7.6	2340	7	AAD48882	Aad48882	Human	ABC
	9	203.6	7.6	2516	6	ABK51682	Abk51682	Human	ABC
	10	194.4	7.3	2035	6	ABK51686	Abk51686	cDNA	enco
	11	193.4	7.2	1915	6	ABK51684	Abk51684	DNA	encod
	12	193.4	7.2	1959	7	AAD48880	Aad48880	Mouse	ABC
	13	193.4	7.2	2258	6	AAD22008	Aad22008	Mouse	sit
	14	193.4	7.2	2354	6	ABK51685	Abk51685	Mouse	ABC
	15	173.2	6.5	1069	6	ABK51687	Abk51687	cDNA	enco
	16	172	6.4	363	6	ABN16253	Abn16253	Human	ORF
c	17	161.4	6.0	5460	6	ABK51683	Abk51683	Human	ABC
	18	155.8	5.8	2525	3	AAZ98625	Aaz98625	Silkworm	
	19	122	4.6	2025	6	ABA94371	Aba94371	Murine	BC
	20	121	4.5	2446	3	AAC37975	Aac37975	Arabidops	
	21	115.4	4.3	1968	6	AAL42412	Aal42412	Human	BCR
	22	115.4	4.3	1968	9	ADC54181	Adc54181	Human	bre
	23	115.4	4.3	1998	6	AAL42414	Aal42414	Human	BCR
	24	115.4	4.3	2027	6	ABK49901	Abk49901	cDNA	enco
	25	115.4	4.3	2053	6	ABK49911	Abk49911	cDNA	enco
	26	115.4	4.3	2247	6	ABA94383	Aba94383	Human	BCR
	27	115.4	4.3	2418	2	AAZ06360	Aaz06360	Breast	Ca
	28	115.4	4.3	2574	4	AAF27724	Aaf27724	Human	tra
	29	115.4	4.3	2574	8	ADA10916	Ada10916	Human	cDN
	30	115.4	4.3	2718	7	ACC80605	Acc80605	Human	ABC
	31	115.4	4.3	2719	3	AAZ94760	Aaz94760	Human	ATP
	32	115.4	4.3	2719	3	AAA27938	Aaa27938	ATP-bindi	
	33	115.4	4.3	2719	6	ABA94369	Aba94369	Human	BCR
	34	115.4	4.3	2883	6	ABZ35528	Abz35528	Human	gen
	35	115	4.3	2352	4	ABL05135	Abl05135	Drosophil	
	36	113.8	4.3	1998	6	AAL42413	Aal42413	Human	BCR
	37	112.6	4.2	4646	7	ADA68676	Ada68676	Spirodela	
	38	109	4.1	2512	9	ADB62671	Adb62671	Human	cDN
	39	106.8	4.0	3201	6	ABV74352	Abv74352	Human	ABC
	40	105.2	3.9	2115	4	ABL07415	Abl07415	Drosophil	
	41	105.2	3.9	2930	3	AAZ94747	Aaz94747	Human	ATP
	42	105.2	3.9	2930	6	ABL63321	Abl63321	Breast	ca
	43	105	3.9	727	4	AAH07859	Aah07859	Human	cDN
	44	105	3.9	2077	4	AAH15008	Aah15008	Human	cDN
c	45	101.4	3.8	447	7	ABX46484	Abx46484	Bovine	ES

ALIGNMENTS

RESULT 1

AAD48883

ID AAD48883 standard; DNA; 2669 BP.

XX

AC AAD48883;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG8 DNA.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 100. .2121
 FT /*tag= a
 FT /product= "hABCG8 protein"
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR P-PSDB; AAE31705.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 13; Page 80; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is human ABCG8 DNA
 XX
 SQ Sequence 2669 BP; 595 A; 768 C; 722 G; 584 T; 0 U; 0 Other;

Query Match 100.0%; Score 2669; DB 7; Length 2669;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGTCCCTGCTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 GTGTCCCTGCTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCT 60

 Qy 61 AAGAGAGCTGCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGAAGGCGGCAGAG 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 AAGAGAGCTGCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGAAGGCGGCAGAG 120

 Qy 121 GAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACCTCGGGCCTCCAGGATAGATTG 180

Db	121		GAGAGAGGGCTGCCGAAAGGGGCCACTCCCAGGATACCTCGGGCCTCCAGGATAGATTG	180
Qy	181		TTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTG	240
Db	181		TTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTG	240
Qy	241		GAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAG	300
Db	241		GAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAG	300
Qy	301		CTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGCCAGAATTCTTGTGAGCTGGGC	360
Db	301		CTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGCCAGAATTCTTGTGAGCTGGGC	360
Qy	361		ATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCA	420
Db	361		ATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCA	420
Qy	421		GTTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATC	480
Db	421		GTTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATC	480
Qy	481		AAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGTGGTGAGGAAGTGT	540
Db	481		AAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGTGGTGAGGAAGTGT	540
Qy	541		GTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTG	600
Db	541		GTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTG	600
Qy	601		GCCTTCATTGCCAGATGCGGCTGCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGG	660
Db	601		GCCTTCATTGCCAGATGCGGCTGCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGG	660
Qy	661		GTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAAC	720
Db	661		GTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAAC	720
Qy	721		ATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGAGCTC	780
Db	721		ATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGAGCTC	780
Qy	781		CTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACA	840
Db	781		CTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACA	840
Qy	841		GCCCCAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATC	900
Db	841		GCCCCAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATC	900
Qy	901		TCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACG	960
Db	901		TCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACG	960
Qy	961		TCTGGCACCCCCATCTACTTAGGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATC	1020

Db	961	TCTGGCACCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATC	1020
Qy	1021	GGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATT	1080
Db	1021	GGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATT	1080
Qy	1081	GACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCC	1140
Db	1081	GACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCC	1140
Qy	1141	CTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGAT	1200
Db	1141	CTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGAT	1200
Qy	1201	CTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCACTAGACACCAACTGCCTCCCG	1260
Db	1201	CTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCACTAGACACCAACTGCCTCCCG	1260
Qy	1261	AGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATT	1320
Db	1261	AGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATT	1320
Qy	1321	TCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATG	1380
Db	1321	TCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATG	1380
Qy	1381	TCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGAT	1440
Db	1381	TCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGAT	1440
Qy	1441	ACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTC	1500
Db	1441	ACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTC	1500
Qy	1501	ATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGAAGTGAAGACGGGCTGTAC	1560
Db	1501	ATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGAAGTGAAGACGGGCTGTAC	1560
Qy	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTAC	1620
Db	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTAC	1620
Qy	1621	ATCATCATCTACGGGATGCCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCC	1680
Db	1621	ATCATCATCTACGGGATGCCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCC	1680
Qy	1681	TTCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTG	1740
Db	1681	TTCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTG	1740
Qy	1741	GCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTAC	1800
Db	1741	GCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTAC	1800
Qy	1801	AACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCC	1860
Db	1801	AACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCC	1860

Qy	1861	GCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTTTGAAGGGCTGATGAAGATTTCAG	1920
Db	1861	GCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTTTGAAGGGCTGATGAAGATTTCAG	1920
Qy	1921	TTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGAT	1980
Db	1921	TTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGAT	1980
Qy	1981	AAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTC	2040
Db	1981	AAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTC	2040
Qy	2041	ATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAG	2100
Db	2041	ATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAG	2100
Qy	2101	AAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGC	2160
Db	2101	AAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGC	2160
Qy	2161	AGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCTTCTGGGGACAGTGAGGACAATGA	2220
Db	2161	AGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCTTCTGGGGACAGTGAGGACAATGA	2220
Qy	2221	CCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAG	2280
Db	2221	CCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAG	2280
Qy	2281	GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTCTGCTCACTGGCAGGAGACTGCGAT	2340
Db	2281	GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTCTGCTCACTGGCAGGAGACTGCGAT	2340
Qy	2341	GACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGA	2400
Db	2341	GACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGA	2400
Qy	2401	TATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAAC TAGGAATGAATTGGGTAG	2460
Db	2401	TATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAAC TAGGAATGAATTGGGTAG	2460
Qy	2461	CTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTG	2520
Db	2461	CTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTG	2520
Qy	2521	GCTTCATCTTCCAGGGGCCCCCACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACC	2580
Db	2521	GCTTCATCTTCCAGGGGCCCCCACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACC	2580
Qy	2581	TAAGATGTACCAGCAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGC	2640
Db	2581	TAAGATGTACCAGCAAGATGCCATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGC	2640
Qy	2641	CAACGTGAACAATTAAAAATGTATTGAGC	2669
Db	2641	CAACGTGAACAATTAAAAATGTATTGAGC	2669

RESULT 2

ABK83218

ID ABK83218 standard; cDNA; 3239 BP.

XX

AC ABK83218;

XX

DT 27-AUG-2002 (first entry)

XX

DE Human transporter and ion channel, TRICH9, Incyte ID 6585710CB1, cDNA.

XX

KW Human; ss; gene; transporter and ion channel; TRICH; transport disorder;

KW neurological disorder; muscle disorder; immunological disorder; cancer;

KW scleroderma; systemic lupus erythematosus; allergy; leukaemia;

KW cell proliferative disorder; cervical cancer; breast cancer;

KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;

KW myotonic dystrophy; catatonia; endocrine disorder; diabetes;

KW Grave's disease; gastrointestinal disorder; Crohn's disease;

KW renal disorder; Good pasture's syndrome; viral infection; cirrhosis;

KW bacterial infection; fungal infection; parasitic infection;

KW protozoal infection; helminthic infection; cardiovascular disorder;

KW atherosclerosis; hepatic disease.

XX

OS Homo sapiens.

XX

PN WO200240541-A2.

XX

PD 23-MAY-2002.

XX

PF 25-OCT-2001; 2001WO-US046055.

XX

PR 27-OCT-2000; 2000US-0243989P.

PR 03-NOV-2000; 2000US-0245904P.

PR 09-NOV-2000; 2000US-0247673P.

PR 17-NOV-2000; 2000US-0249661P.

PR 20-NOV-2000; 2000US-0252232P.

PR 01-DEC-2000; 2000US-0250790P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Tang YT, Yue H, Nguyen DB, Hafalia AJA, Elliott VS, Lu Y;

PI Walia NK, Yao MG, Baughn MR, Gandhi AR, Ding L, Sanjanwala M;

PI Ramkumar J, Arvizu C, Gietzen KJ, Lal PG, Azimzai Y, Khan FA;

PI Thangavelu K, Thornton M, Lu DAM, Tribouley CM, Warren BA, Ison CH;

PI Das D, Raumann BE, Policky JL, Kearney L;

XX

DR WPI; 2002-463570/49.

DR P-PSDB; ABG61539.

XX

PT New transporters and ion channels (TRICH) polypeptides, useful for

PT diagnosing, preventing, and treating disorders associated with an

PT abnormal expression or activity of TRICH, e.g. immunological, muscular or

PT renal disorders.

XX

PS Claim 5; Page 167-168; 178pp; English.

XX

CC The invention relates to human transporters and ion channels (TRICH)

polypeptides, a naturally occurring amino acid sequence 90 % identical to TRICH, a biologically active fragment of TRICH or an immunogenic fragment of TRICH. Also included are an isolated polynucleotide encoding TRICH, a recombinant polynucleotide comprising a promoter sequence operably linked to the TRICH polynucleotide, a cell transformed with the recombinant polynucleotide, a transgenic organism comprising the recombinant polynucleotide, an isolated antibody that binds specifically to TRICH, and screening for compounds which bind to TRICH, modulate TRICH, modulate TRICH expression or are ant/agonists of TRICH. The polypeptides are useful for diagnosing, treating, and preventing transport, neurological, muscle, immunological disorders (e.g. scleroderma, systemic lupus erythematosus, allergies), cell proliferative disorders such as cancers (e.g. leukaemia, cervical or breast cancers), neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's disease), muscular disorders (e.g. myotonic dystrophy, catatonia), endocrine disorders (e.g. diabetes, Grave's disease), gastrointestinal disorders (e.g. Crohn's disease), renal disorders (e.g. Good pasture's syndrome), viral, bacterial, fungal, parasitic, protozoal and helminthic infections, cardiovascular disorders (e.g. atherosclerosis), or hepatic diseases (e.g. cirrhosis) and many other diseases and disorders detailed in the specification. They can also be used in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels. TRICH or its fragments may also be used in screening for compounds that specifically bind to and modulate the activity of TRICH. The polynucleotides can be used to create knock-in humanised animals or transgenic animals to model human disease. The present sequence encodes a TRICH protein

Sequence 3239 BP; 784 A; 822 C; 796 G; 837 T; 0 U; 0 Other;

Query Match 63.0%; Score 1680.6; DB 6; Length 3239;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1683; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

[illegible]

Db

Qy

Db

Qy

Db

Qy

Db

QY

Db

Qy

Db

Qy	1343	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	1402
Db	372	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	431
Qy	1403	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA	1462
Db	432	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA	491
Qy	1463	TCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGA	1522
Db	492	TCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGA	551
Qy	1523	GGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG	1582
Db	552	GGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG	611
Qy	1583	CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCCCA	1642
Db	612	CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCCCA	671
Qy	1643	CCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGG	1702
Db	672	CCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGG	731
Qy	1703	TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA	1762
Db	732	TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA	791
Qy	1763	CCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACTCCTTCTACCTCGCCGGGG	1822
Db	792	CCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACTCCTTCTACCTCGCCGGGG	851
Qy	1823	GCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCT	1882
Db	852	GCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCT	911
Qy	1883	TCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTAGTTTCAGCAGAAGAACTTATAAAA	1942
Db	912	TCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTAGTTTCAGCAGAAGAACTTATAAAA	971
Qy	1943	TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC	2002
Db	972	TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC	1031
Qy	2003	TGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA	2062
Db	1032	TGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA	1091
Qy	2063	TGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGAT	2122
Db	1092	TGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGAT	1151
Qy	2123	TCACGCCAGACGTCTGCCCCTGGTGGGGGACCTGAGCAGACCTTCAACTGCACTCCCT	2182
Db	1152	TCACGCCAGACGTCTGCCCCTGGTGGGGGACCTGAGCAGACCTTCAACTGCACTCCCT	1211

Qy	2183	CCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACAT	2242
Db	1212	CCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACAT	1271
Qy	2243	CCGCCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAG	2302
Db	1272	CCGCCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAG	1331
Qy	2303	TCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTC	2362
Db	1332	TCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTC	1391
Qy	2363	GGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATATAGGCAACTC	2422
Db	1392	GGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATATAGGCAACTC	1451
Qy	2423	GATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTG	2482
Db	1452	GATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTG	1511
Qy	2483	GAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTCCAGGGGCCCCA	2542
Db	1512	GAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTCCAGGGGCCCCA	1571
Qy	2543	CACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATGCC	2602
Db	1572	CACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATGCC	1631
Qy	2603	ATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAATTAAAAATGT	2662
Db	1632	ATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAATTAAAAATGT	1691
Qy	2663	ATTGAGC	2669
Db	1692	ATTGAGC	1698

RESULT 3

ABN90022

ID ABN90022 standard; cDNA; 2564 BP.

XX

AC ABN90022;

XX

DT 16-AUG-2002 (first entry)

XX

DE Mouse clone IMX3_67 extended sequence.

XX

KW Mouse; antiinflammatory; gene therapy; ileitis; DST; ss; TOGA;

KW digital sequence tag; total gene expression analysis.

XX

OS Mus musculus.

XX

PN WO200231114-A2.

XX

PD 18-APR-2002.

XX

PF 11-OCT-2001; 2001WO-US032091.


```

XX 11-OCT-2000; 2000US-0239483P.
XX
PA (DIGI-) DIGITAL GENE TECHNOLOGIES INC.
XX
PI Viney JL, Sims JE, Dubose RF, Baum PR, Hasel KW, Hilbush BS;
XX
DR WPI; 2002-426279/45.
XX
PT New isolated nucleic acid molecules that are associated with ileitis, for
PT preventing, treating, modulating and diagnosing ileitis in a mammalian
PT subject.
XX
PS Claim 1; Page 266-268; 273pp; English.
XX
CC The invention relates to a novel isolated nucleic acid molecule
CC comprising a polynucleotide having one of 90 polynucleotide sequences,
CC given in the specification. The polynucleotides of the invention have
CC antiinflammatory activity, and may have a use in gene therapy. The
CC polynucleotide or a polypeptide encoded by it is used for preventing,
CC treating, modulating or ameliorating a medical condition such as ileitis.
CC The polypeptide or polynucleotide is also useful for manufacturing a
CC medicament for treating ileitis. The sequence represents a an extended
CC cDNA digital sequence tag obtained from a mouse clone by the TOGA (total
CC gene expression analysis) method
XX
SQ Sequence 2564 BP; 623 A; 722 C; 638 G; 581 T; 0 U; 0 Other;

Query Match          56.2%; Score 1499; DB 6; Length 2564;
Best Local Similarity 77.2%; Pred. No. 0;
Matches 1933; Conservative 0; Mismatches 520; Indels 52; Gaps 7;

Qy      99 CATGCCCGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATAC 158
      ||||| | ||| | ||| ||| ||| ||| ||| | | ||||| |
Db      34 CATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGC 93

Qy     159 CTC---GGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCAC 215
      || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      94 TTCGCAGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCAC 153

Qy     216 CTACAGTGGCCAGCCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGC 275
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     154 CTACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGC 213

Qy     276 CTCTCAGGTCCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAG 335
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     214 CTCTCAGGTGCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAG 273

Qy     336 CTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCA 395
      | |||| | || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db     274 CAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACA 333

Qy     396 GATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCAC 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     334 GATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCAC 393

Qy     456 TGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAG 515

```

Db	394	AGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAG	453
Qy	516	CTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCC	575
Db	454	TACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCC	513
Qy	576	CAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTT	635
Db	514	CAACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTT	573
Qy	636	CTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCA	695
Db	574	CTCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCA	633
Qy	696	GTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGTGAGCGCAG	755
Db	634	GTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGTGAGCGCCG	693
Qy	756	GAGAGTCAGCATTGGGGTGAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACC	815
Db	694	ACGAGTGAGCATTGGGGTGAGCTCCTGTGGAACCCAGGAATCCTCATTCCTGGATGAACC	753
Qy	816	CACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGC	875
Db	754	CACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGC	813
Qy	876	CAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCT	935
Db	814	CAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCT	873
Qy	936	GTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCA	995
Db	874	ATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCATCTACCTGGGGGCGGCGCAGCA	933
Qy	996	CATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGA	1055
Db	934	AATGGTGAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCTTGCGGA	993
Qy	1056	CTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAG	1115
Db	994	CTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGT	1053
Qy	1116	GGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTT	1175
Db	1054	GGAGAAGGCACAGTCTCTTGACAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTT	1113
Qy	1176	TCTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGAC	1235
Db	1114	TCTGTGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTGACCCT	1173
Qy	1236	CCCACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTT	1295
Db	1174	CACACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCGGGATGATAGAGCAGTT	1230
Qy	1296	TACGACGCTGATCCGTGCTCAGATTTCCAACGACTTCCGAGACCTGCCACCCCTCCTCAT	1355

Db 1231 TTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCAT 1290
 Qy 1356 CCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTGGCCATGG 1415
 ||||| |||| ||||| ||||| | | ||| ||||| || | |||||
 Db 1291 TCATGGGTTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGG 1350
 Qy 1416 GAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCAT 1475
 | || ||||| ||||| ||||| ||||| | ||||| || || |||||
 Db 1351 GGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCAT 1410
 Qy 1476 CCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTA 1535
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 Db 1411 TCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTA 1470
 Qy 1536 CTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGG 1595
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 Db 1471 CTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCTTATTTCTTTGCCAAGATCCTAGG 1530
 Qy 1596 GGAGCTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCCACCTACTGGCTGGC 1655
 || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 Db 1531 AGAATTGCCGAGCACTGTGCCTACGTATCATCTACGCGATGCCCATCTACTGGCTGAC 1590
 Qy 1656 CAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGT 1715
 ||||| |||| | | ||| ||||| || ||||| ||||| ||||| |||||
 Db 1591 AAACCTGCGGCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTGTGGTGGT 1650
 Qy 1716 CTTCTGTTGCAGGATTATGGCCCTGGCCGCCCGGCCCTGCTCCCCACCTTCCACATGGC 1775
 ||||| ||||| ||||| ||||| ||| | ||| ||| ||||| ||||| ||
 Db 1651 CTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTC 1710
 Qy 1776 CTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGCTTCATGATAAA 1835
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1711 CTCCTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAA 1770
 Qy 1836 CTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTTCCTGCGGTGGTG 1895
 |||| || ||||| ||||| || ||||| |||| |||| ||||| |||||
 Db 1771 CTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTTCCTCCGGTGGTG 1830
 Qy 1896 TTTTGAAGGGCTGATGAAGATTCAAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAA 1955
 || | ||||| ||||| || | || | || | | | |||| ||
 Db 1831 CTTCTCGGTGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAA 1890
 Qy 1956 CCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCC 2015
 | |||| || | || ||||| | || ||||| |||| |||| | ||
 Db 1891 CTTACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCC 1950
 Qy 2016 TCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCTGTACTA 2075
 ||||| || ||||| ||||| |||| |||| ||||| ||||| ||||| |||||
 Db 1951 ACTCTATGCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTA 2010
 Qy 2076 CGTGTCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGATTACGCCAGACGT 2135
 | |||| | | ||||| |||| ||||| ||||| || | | |||||
 Db 2011 TCTATCCTTGAAGCTCATCAAACAGAAAGTCAATTCAAGACTGGTGATACTCAGCCTTGCT 2070
 Qy 2136 CTGCCCCTGGTGGGGGACCTGAGCAGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCC 2195
 || | | || | | || | | || | | || | | || | | ||
 Db 2071 CTCCTGCGCG-----GACCCTTTTCCCGGGGCTGGCCACCCAGGAGGAGCC 2118

Qy	2196	TTCTGGGGACAGTGAGGACAATGACCCCTACAGATGCTCAGCTACATCCGGCCCCAGGGTG	2255
Db	2119	GGACTGGGGACAAGGCTCACACAGATCTCTCAG-----GCAGCAGCCACCTCTTAGTG	2171
Qy	2256	CTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAGTCGAAAGGGATT	2315
Db	2172	CTGCAGTGGCACAGATCAGCCACAGGATGGCAGTAGAATAAAGACAGTTGAGAGGTGTTT	2231
Qy	2316	CTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTCGGTGGCACCTACA	2375
Db	2232	CTGCTCCCAGGCCAGGTTTGTAAATGGGAGAGAGAGAA-----ACCAGGT	2276
Qy	2376	ACGTTTGCTAATTTATTTTCCTTTTGATATGCATTTATATAGGCAACTCGATATAGGATGGG	2435
Db	2277	ACGTTGCTCATGCATTT-----TATATCTTTAAATAAAACAACCCAGTATGGAATGGG	2328
Qy	2436	AGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGA	2495
Db	2329	AACCAATTATATATGAATTGAGTAGCTAGGCTATGCAGAAATTTCTGGAATCCTGAGAGG	2388
Qy	2496	ACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCCAGGGGCCCCACACTCCGTGGTGA	2555
Db	2389	ATAGTGGGTTTATAGCAAAGTGTTTAACTTTCTCTTCTACCATTCTCACAC---TGTTAA	2444
Qy	2556	GCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATG	2600
Db	2445	GCCACTCCCAATACAAAGGGCGACCTAAACAAACTAGCAAAATG	2489

```
XX PF      20-NOV-2001; 2001WO-US043823.
XX
PR      20-NOV-2000; 2000US-0252235P.
PR      28-NOV-2000; 2000US-0253645P.
XX
PA      (TULA-) TULARIK INC.
PA      (TEXA ) UNIV TEXAS SYSTEM.
XX
PI      Hobbs HH,   Shan B,   Barnes R,   Tian H;
XX
DR      WPI; 2003-058548/05.
DR      P-PSDB; AAE31703.
XX
PT      New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
PT      related disorders e.g. sitosterolemia, hypercholesterolemia,
PT      hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT      nutritional deficiencies.
XX
PS      Claim 13; Page 75; 94pp; English.
XX
CC      The invention relates to ATP-binding cassette (ABC) family cholesterol
CC      transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC      provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC      as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC      are useful for treating or preventing sterol-related disorders such as
CC      sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC      deficiency, atherosclerosis and nutritional deficiencies. They are also
CC      useful in gene therapy. The present sequence is mouse ABCG8 DNA
XX
SQ      Sequence 2019 BP; 444 A; 598 C; 510 G; 467 T; 0 U; 0 Other;

Query Match          53.6%; Score 1430; DB 7; Length 2019;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 1659; Conservative    0; Mismatches 360; Indels    3; Gaps    1;

Qy      100 ATGGCCGGGAAGGCCGAGAGGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACC 159
        ||||| | ||| | ||| |||| |||| ||| ||| | | ||||| |
Db      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60

Qy      160 TCGGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTAC 219
        |||||||||||||| || |||||||||| |||||||||| |||||||||| |||
Db      61 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120

Qy      220 AGTGGCCAGCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCT 279
        ||||| ||| ||||||| |||||||||| |||| |||||||||| ||| |||||
Db      121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180

Qy      280 CAGGTCCCTTGTTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGC 339
        ||||| |||||||||||||||||||||| ||||||| |||| ||| ||
Db      181 CAGGTGCCTTGTTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240

Qy      340 CAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATG 399
        || | || |||||||||| || |||||||||| |||| |||||
Db      241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300

Qy      400 CTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGC 459
```

Db	301	CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC	360
Qy	460	CGAGGTACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCG	519
Db	361	AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG	420
Qy	520	CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAAC	579
Db	421	CCTCAGCTGGTGAGGAAGTGCCTTGCATGTGCGGCAGCATGACCAACTGCTGCCAAC	480
Qy	580	TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCC	639
Db	481	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC	540
Qy	640	CAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGC	699
Db	541	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Qy	700	GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGA	759
Db	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGTGAGCGCCGACGA	660
Qy	760	GTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC	819
Db	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Qy	820	TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAA	879
Db	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCTGGCCAAAG	780
Qy	880	GGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTT	939
Db	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Qy	940	GATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACATG	999
Db	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Qy	1000	GTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTC	1059
Db	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCTGCGGACTTC	960
Qy	1060	TATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAG	1119
Db	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Qy	1120	AAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCCTGACTTAGATGACTTTCTA	1179
Db	1021	AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Qy	1180	TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAGCAGCGTGACCCCA	1239
Db	1081	TGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Qy	1240	CTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG	1299

Db	1141	CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTC	1197
Qy	1300	ACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCACCCTCCTCATCCAT	1359
Db	1198	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT	1257
Qy	1360	GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGC	1419
Db	1258	GGGTGCGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1317
Qy	1420	ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCT	1479
Db	1318	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1377
Qy	1480	TTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT	1539
Db	1378	TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT	1437
Qy	1540	GAAGTGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG	1599
Db	1438	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA	1497
Qy	1600	CTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAAC	1659
Db	1498	TTGCCGGAGCACTGTGCCTACGTCACTCTACGCGATGCCCATCTACTGGCTGACAAAC	1557
Qy	1660	CTGAGGCCAGGCCTCCAGCCCTTCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC	1719
Db	1558	CTGCGGCCCGTGCCCTGAGCTCTTCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC	1617
Qy	1720	TGTTGCAGGATTATGGCCCTGGCCGCCCGGGCCCTGCTCCCCACCTTCCACATGGCCTCC	1779
Db	1618	TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCTCTCC	1677
Qy	1780	TTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG	1839
Db	1678	TTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG	1737
Qy	1840	AGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTTTT	1899
Db	1738	GACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCTTCCTCCGGTGGTGCTTC	1797
Qy	1900	GAAGGGCTGATGAAGATTCAAGTTCAGCAGAGAAGAACTTATAAAATGCCTCTCGGGAACCTC	1959
Db	1798	TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC	1857
Qy	1960	ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCCTCTC	2019
Db	1858	ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCCACTC	1917
Qy	2020	TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCTGTACTACGTG	2079
Db	1918	TATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA	1977
Qy	2080	TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA	2121
Db	1978	TCCTTGAAGCTCATCAAACAGAAAGTCAATTCAAGACTGGTGA	2019

RESULT 5

AAH98911

ID AAH98911 standard; cDNA; 580 BP.

XX

AC AAH98911;

XX

DT 12-OCT-2001 (first entry)

XX

DE Arabidopsis EST-derived coding sequence SEQ ID NO: 768.

XX

KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;

KW gene therapy; nutrition; ss.

XX

OS Arabidopsis thaliana.

XX

PN WO200154477-A2.

XX

PD 02-AUG-2001.

XX

PF 25-JAN-2001; 2001WO-US002687.

XX

PR 25-JAN-2000; 2000US-00491404.

PR 17-JUL-2000; 2000US-00617746.

PR 03-AUG-2000; 2000US-00631451.

PR 15-SEP-2000; 2000US-00663870.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX

DR WPI; 2001-476164/51.

DR P-PSDB; AAM24252.

XX

PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.

XX

PS Claim 1; Page 664; 1275pp; English.

XX

CC The present invention provides the protein and coding sequences of novel

CC proteins from a variety of organisms, including human, dog, cat, horse,

CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)

CC from the organism of interest. They can be used in diagnostics,

CC forensics, gene mapping, identification of mutations, to assess

CC biodiversity and for nutritional purposes. The present sequence is a cDNA

CC of the invention

XX

SQ Sequence 580 BP; 146 A; 154 C; 116 G; 164 T; 0 U; 0 Other;

Query Match 10.9%; Score 291.6; DB 4; Length 580;

Best Local Similarity 97.1%; Pred. No. 9.9e-68;

Matches 297; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

[illegible]

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR P-PSDB; AAU98984.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.

XX

PS Claim 38; Page 36-37; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the human ABCG5 gene located on chromosome 2p21.
CC This sequence encodes the human ABCG5 protein of the invention

XX

SQ Sequence 1920 BP; 440 A; 503 C; 486 G; 491 T; 0 U; 0 Other;

Query Match 7.6%; Score 203.6; DB 6; Length 1920;
Best Local Similarity 54.4%; Pred. No. 8.2e-44;
Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

Qy 335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
|| | ||| | | || | | | | | | | | | | | | | | |
Db 143 GCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGC 202

Qy 395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
|||| | | ||| ||| ||||| | | ||| | | | | | | | |
Db 203 AGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGT 262

Qy 455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
| | | | | | | | | | | | | | | | | | | | | |
Db 263 CCGGGAGGCTGGGGCGCGGGGACCTTCTGGGGAGGTGTATGTGAACGGCCGGGGCGC 322

Qy 515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCC 574
| ||| | | | | | | | | | | | | | | | | | |

XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 107..2062
 FT /*tag= a
 FT /product= "Human SSG protein"
 XX
 PN WO200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001WO-US012758.
 XX
 PR 18-APR-2000; 2000US-0198465P.
 PR 15-MAY-2000; 2000US-0204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 XX
 DR WPI; 2002-017598/02.
 DR P-PSDB; AAE13290.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
 PT useful for screening a compound that increases the level of expression or
 PT activity of SSG polypeptide for treating sterol-related disorder.
 XX
 PS Claim 8; Fig 8; 105pp; English.
 XX
 CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
 CC identifying a compound useful in the treatment or prevention of a sterol-
 CC related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations, and
 CC for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG DNA. Human SSG is located on chromosome
 CC 2p21
 XX
 SQ Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

Query Match 7.6%; Score 203.6; DB 6; Length 2340;
 Best Local Similarity 54.4%; Pred. No. 9.1e-44;
 Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

Qy 335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
 || | ||| | | || | || | | || | ||| | ||| || ||||
 Db 285 GCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGGGC 344

Qy 395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
 ||| | | | | | | | | | | | | | | | | |
 Db 345 AGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGT 404
 Qy 455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
 | | | | | | | | | | | | | | | | | |
 Db 405 CCGGGAGGCTGGGGCGCGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGCGC 464
 Qy 515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCC 574
 | | | | | | | | | | | | | | | | | |
 Db 465 TGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGA 524
 Qy 575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
 | | | | | | | | | | | | | | | | | |
 Db 525 GCAGCCTCACCGTGCAGGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCA 584
 Qy 635 TCTCCCAGGCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
 | | | | | | | | | | | | | | | | | |
 Db 585 ATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCC 641
 Qy 695 AGTGCCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA 754
 | | | | | | | | | | | | | | | | | |
 Db 642 ATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGC 701
 Qy 755 GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
 | | | | | | | | | | | | | | | | | |
 Db 702 GCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGC 761
 Qy 815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG 874
 | | | | | | | | | | | | | | | | | |
 Db 762 CAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCTGTCCTCCTGGTGGAACTGG 821
 Qy 875 CCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
 | | | | | | | | | | | | | | | | | |
 Db 822 CTCGCAGGAACCGAATTGTGGTTCTACCATTCACCAGCCCCGTTCTGAGCTTTTTTCAGC 881
 Qy 935 TGTTTGATCTGGTCCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGC 994
 | | | | | | | | | | | | | | | | | |
 Db 882 TCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGG 941
 Qy 995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTG 1054
 | | | | | | | | | | | | | | | | | |
 Db 942 AAATGCTTGATTTCTTCAATGACTGCGGTTACCCTGTCTCTGAACATTCAAACCCTTTTG 1001
 Qy 1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
 | | | | | | | | | | | | | | | | | |
 Db 1002 ACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCT 1061
 Qy 1115 GGGAGAAGGCTCAG 1128
 | | | | | |
 Db 1062 CCAAGAGAGTCCAG 1075

RESULT 8

AAD48882

ID AAD48882 standard; DNA; 2340 BP.

XX
 AC AAD48882;
 XX
 DT 24-MAR-2003 (first entry)
 XX
 DE Human ABCG5 DNA.
 XX
 KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 107. .2062
 FT /*tag= a
 FT /product= "hABCG5 protein"
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR P-PSDB; AAE31704.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 11; Page 77; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is human ABCG5 DNA
 XX
 SQ Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

Query Match 7.6%; Score 203.6; DB 7; Length 2340;
 Best Local Similarity 54.4%; Pred. No. 9.1e-44;

Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

```
Qy      335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      285 GCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGC 344

Qy      395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      345 AGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGT 404

Qy      455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      405 CCGGGAGGCTGGGGCGCGCGGGGACCTTCTCTGGGGGAGGTGTATGTGAACGGCCGGGCGC 464

Qy      515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCC 574
          | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      465 TGCGCCGGGAGCAGTTCCAGGACTGCTTCTCTACGTCTCTGCAGAGCGACACCCTGCTGA 524

Qy      575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      525 GCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCCGGGCA 584

Qy      635 TCTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
          || || || || || || || || || || || || || || || ||
Db      585 ATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCC 641

Qy      695 AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA 754
          | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      642 ATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGC 701

Qy      755 GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
          | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      702 GCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGC 761

Qy      815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG 874
          | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      762 CAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGAACCTGG 821

Qy      875 CCAAAGGCAACCGGCTGGTGTCTATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
          | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      822 CTCGCAGGAACCGAATTGTGGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGC 881

Qy      935 TGTTTGATCTGGTCTCTCTGATGACGTCTGGCACCCCCATCTACTTAGGGGCGGGCCAGC 994
          | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      882 TCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGG 941

Qy      995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCTCGCTACAGCAATCCTGCTG 1054
          | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      942 AAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCTCTGAACATTCAAACCCTTTTG 1001

Qy      1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1002 ACTTCTATATGGACCTGACGTCAGTGATACCCAAAGCAAGGAACGGGAAATAGAAACCT 1061

Qy      1115 GGGAGAAGGCTCAG 1128
          ||| ||| |||
Db      1062 CCAAGAGAGTCCAG 1075
```

RESULT 9

ABK51682

ID ABK51682 standard; cDNA; 2516 BP.

XX

AC ABK51682;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 cDNA sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW chromosome 2p21; ss.

XX

OS Homo sapiens.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 37-38; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's

CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the cDNA sequence of human ABCG5 gene located on
CC chromosome 2p21

XX

SQ Sequence 2516 BP; 601 A; 631 C; 636 G; 648 T; 0 U; 0 Other;

Query Match 7.6%; Score 203.6; DB 6; Length 2516;
Best Local Similarity 54.4%; Pred. No. 9.4e-44;
Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

```
Qy      335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
      || | ||| | |      || | | | | | | | | ||| | ||| || ||||
Db      319 GCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGC 378

Qy      395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
      |||| | | ||| ||| ||||| | | ||| | | | |||| | | | ||
Db      379 AGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGT 438

Qy      455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
      | || | | ||| ||| || | | | | | | | | | | | | | |
Db      439 CCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGC 498

Qy      515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACAGTGCGCCAGCACAAACCAGCTGCTCC 574
      | ||| | | | | | | | | | | | | | | | | | | | | | |
Db      499 TGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCTGA 558

Qy      575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
      || | | | | |||| | || | | | | | | | | | | | | | |
Db      559 GCAGCCTCACCGTGCGCGAGACGCTGCACTACACGCGCTGCTGGCCATCCGCCGCGGCA 618

Qy      635 TCTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
      || || | | | | | | | | | | | | | | | | | | | |
Db      619 ATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTGATGGCAGAGCTGAGTCTGAGCC 675

Qy      695 AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGTGAGCGCA 754
      | || | | | | | | | | | | | | | | | | | | | | |
Db      676 ATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGC 735

Qy      755 GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
      | | ||| ||| | | |||| | | | | | | | | | | | | | |
Db      736 GCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGC 795

Qy      815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCAGGCTGG 874
      | ||| | | | | | | | | | | | | | | | | | | | | | |
Db      796 CAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTGTCCTCCTGGTGGAACTGG 855

Qy      875 CCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
      | | | |||| | ||| | | || | | ||||| | | |||| | || | |
Db      856 CTCGCAGGAACCGAATTGTGGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGC 915

Qy      935 TGTTTGATCTGGTCCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGC 994
      | |||| | | | | | | | | | | | | | | | | | | | |
Db      916 TCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGG 975

Qy      995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTG 1054
      | ||| | | | |||| | | ||| |||| | |||| | | ||| ||
```

Db 976 AAATGCTTGATTCTTCAATGACTGCGGTTACCCCTTGTCCTGAACATTCAAACCCTTTTG 1035

QY 1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
 ||||| ||||| | || | ||| || | ||| | | |||

Db 1036 ACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCT 1095

QY 1115 GGGAGAAGGCTCAG 1128
 ||| | |||

Db 1096 CCAAGAGAGTCCAG 1109

RESULT 10

ABK51686

ID ABK51686 standard; cDNA; 2035 BP.

XX

AC ABK51686;

XX

DT 07-AUG-2003 (revised)

DT 30-JUL-2002 (first entry)

XX

DE cDNA encoding rat ABCG5 protein.

XX

KW Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; ss;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

XX

OS Rattus sp.

XX

FH Key Location/Qualifiers

FT CDS 8. .1965

FT /*tag= a

FT /product= "Rat ABCG5 protein"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR P-PSDB; AAU96986.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 45-46; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a

CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the rat ABCG5 protein of the invention. (Updated on
 CC 07-AUG-2003 to correct OS field.)

XX

SQ Sequence 2035 BP; 481 A; 533 C; 537 G; 484 T; 0 U; 0 Other;

Query Match 7.3%; Score 194.4; DB 6; Length 2035;
 Best Local Similarity 54.0%; Pred. No. 2.5e-41;
 Matches 421; Conservative 0; Mismatches 356; Indels 3; Gaps 1;

Qy	360	CATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTC	419
Db	214	CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGACCATGTGCATCTTAGGTAGCTC	273
Qy	420	AGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGAT	479
Db	274	AGGCTCAGGGAAAACCACGCTGCTGGACGCCATCTCTGGGAGGCTGCGGCGCACAGGGAC	333
Qy	480	CAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAGTG	539
Db	334	CTTGGAAGGGGAAGTGTTTGTGAACGGCTGCGAGCTGCGCAGGGACCAAGACTG	393
Qy	540	TGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTT	599
Db	394	CGTCTCCTACCTCCTGCAGAGCGATGTCTTTCTGAGCAGCCTCACGGTGCGGGAGACGCT	453
Qy	600	GGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCAGGCCCAGCGTGACAAAAG	659
Db	454	GAGATACACGGC---GATGCTGGCTCTCCGCAGCAGCTCCGCGGACTTCTACGACAAGAA	510
Qy	660	GGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAA	719
Db	511	GGTAGAGGCAGTCCTGACAGAGCTGAGTCTGAGCCACGTGGCAGACCAAATGATCGGCAA	570
Qy	720	CATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGACGCT	779
Db	571	CTATAATTTTGGGGGGATTTCAGTGCGGAGCGGCGCCGAGTGTCCATCGCAGCCCAACT	630
Qy	780	CCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCAC	839
Db	631	CCTTCAGGACCCCAAGGTCATGATGCTTGACGAGCCAACACAGGACTGGACTGCATGAC	690

Qy 840 AGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCAT 899
 || | | | | | | | | | | | | | | | | | | | |
 Db 691 TGCAAATCATATCGTCCTCCTCTTGGTTCGAGCTGGCTCGCAGGAACCGCATTGTAATTGT 750

Qy 900 CTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGAC 959
 | | | | | | | | | | | | | | | | | | | | | |
 Db 751 CACCATCCACCAGCCTCGCTCTGAGCTCTTCCACCACTTCGACAAAATTGCCATTCTGAC 810

Qy 960 GTCTGGCACCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCAT 1019
 | | | | | | | | | | | | | | | | | | | | | |
 Db 811 TTACGGAGAGTTGGTGTCTGTGGCACGCCAGAGGAGATGCTCGGCTTCTTCAATAACTG 870

Qy 1020 CGGCTACCCCTGTCTCGCTACAGCAATCTGCTGACTTCTATGTGGACCTGACCAGCAT 1079
 || | | | | | | | | | | | | | | | | | | | |
 Db 871 TGGTTACCCCTGTCTGAACATCCAATCCCTTTGATTCTACATGGACTTGACATCGGT 930

Qy 1080 TGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGC 1139
 |||| | | | | | | | | | | | | | | | | | | | |
 Db 931 GGACACCCAAAGCAGAGAGCGAGAGATAGAGACGTACAAGCGAGTCCAGATGCTGGAATC 990

RESULT 11

ABK51684

ID ABK51684 standard; DNA; 1915 BP.

XX

AC ABK51684;

XX

DT 30-JUL-2002 (first entry)

XX

DE DNA encoding mouse ABCG5 protein.

XX

KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ds.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT CDS 1. .1915

FT /*tag= a

FT /partial

FT /product= "Mouse ABCG5 protein"

FT /transl_except= (pos: 1912. .1915, aa: LGIVIFKVRDY LISR)

FT /note= "This sequence lacks a stop codon"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX
PI Patel SB, Dean M;
XX
DR WPI; 2002-416483/44.
DR P-PSDB; AAU96985.
XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.
XX
PS Example 3; Page 42-43; 66pp; English.

XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence encodes the mouse ABCG5 protein of the invention

XX
SQ Sequence 1915 BP; 453 A; 502 C; 484 G; 476 T; 0 U; 0 Other;

Query Match 7.2%; Score 193.4; DB 6; Length 1915;
Best Local Similarity 53.4%; Pred. No. 4.5e-41;
Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

Qy	335	GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC	394
Db	182	GCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCC	241
Qy	395	AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA	454
Db	242	AGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCT	301
Qy	455	CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA	514
Db	302	CCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGC	361
Qy	515	GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCC	574
Db	362	TGCGCAGGGACCACTTCCAAGACTGCTTCTCCTACGTCTGCAGAGCGACGTTTTTCTGA	421
Qy	575	CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCAGATGCGGCTGCCCGAAGCT	634

FH Key Location/Qualifiers
FT CDS 1. .1591
FT /*tag= a
FT /product= "mABCG5 protein"

XX
PN WO200281691-A2.
XX
PD 17-OCT-2002.
XX
PF 20-NOV-2001; 2001WO-US043823.
XX
PR 20-NOV-2000; 2000US-0252235P.
PR 28-NOV-2000; 2000US-0253645P.
XX
PA (TULA-) TULARIK INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tian H;

XX
DR WPI; 2003-058548/05.
DR P-PSDB; AAE31702.
XX
PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
PT related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies.

XX
PS Claim 11; Page 73; 94pp; English.

XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is mouse ABCG5 DNA

XX
SQ Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 U; 0 Other;

Query Match 7.2%; Score 193.4; DB 7; Length 1959;
Best Local Similarity 53.4%; Pred. No. 4.6e-41;
Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

Qy 335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
|| ||||| | || | | | | || | | ||||| |
Db 182 GCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCC 241

Qy 395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
|||| || |||| |||| ||||| | |||| || | |||| || | |||
Db 242 AGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCT 301

Qy 455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
| || | | || || || | | || | || | |||||
Db 302 CCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGC 361

Qy 515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCC 574

Db	362	TGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGA	421
Qy	575	CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCAGAACCT	634
Db	422	GCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGC---GATGCTGGCCCTCTGCCGCA	478
Qy	635	TCTCCCAGGCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC	694
Db	479	GCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCC	538
Qy	695	AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA	754
Db	539	ACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGC	598
Qy	755	GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC	814
Db	599	GCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGC	658
Qy	815	CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG	874
Db	659	CAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCTTCTCTTGGCTGAGCTGG	718
Qy	875	CCAAAGGCAACCGGCTGGTGTCTCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC	934
Db	719	CTCGCAGGGACCGAATTGTGATTGTCAACCATCCACCAGCCTCGCTCTGAGCTCTTCCAAC	778
Qy	935	TGTTTGATCTGGTCCTCCTGATGACGCTTGGCACCCCATCTACTTAGGGGCGGCCCAGC	994
Db	779	ACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCAGAGG	838
Qy	995	ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTG	1054
Db	839	AGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCCTGAACATTCCAATCCCTTTG	898
Qy	1055	ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA	1114
Db	899	ATTTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGT	958
Qy	1115	GGGAGAAGGCTCAGTCACTCGCA	1137
Db	959	ACAAGCGAGTACAGATGCTGGAA	981

RESULT 13

AAD22008

ID AAD22008 standard; DNA; 2258 BP.

XX

AC AAD22008;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG).

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.
XX
OS Mus sp.

XX
FH Key Location/Qualifiers
FT CDS 47..2005
FT /*tag= a
FT /product= "Mouse SSG protein"

XX
PN WO200179272-A2.

XX
PD 25-OCT-2001.

XX
PF 18-APR-2001; 2001WO-US012758.

XX
PR 18-APR-2000; 2000US-0198465P.
PR 15-MAY-2000; 2000US-0204234P.

XX
PA (TULA-) TULARIK INC.

XX
PI Tian H, Schultz J, Shan B;

XX
DR WPI; 2002-017598/02.
DR P-PSDB; AAE13289.

XX
PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
PT useful for screening a compound that increases the level of expression or
PT activity of SSG polypeptide for treating sterol-related disorder.

XX
PS Claim 8; Fig 7; 105pp; English.

XX
CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
CC identifying a compound useful in the treatment or prevention of a sterol-
CC related disorder, including sitosterolaemia, hyperlipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolaemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolaemia and other
CC cardiovascular disorders, for forensics and paternity determinations, and
CC for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG DNA. Mouse SSG is located on chromosome 17

XX
SQ Sequence 2258 BP; 549 A; 579 C; 567 G; 563 T; 0 U; 0 Other;

Query Match 7.2%; Score 193.4; DB 6; Length 2258;
Best Local Similarity 53.4%; Pred. No. 4.9e-41;
Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

Qy 335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
|| ||||| | || | || | | || | |||||
Db 228 GCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCC 287

Qy		395 AGATGCTTGGCCATCATAGGGACGTCAAGTTGTGGGAGAGCCTTCCTTGCTAGATGTGATCA	454
Db		288 AGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCT	347
Qy		455 CTGGCCCAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCCA	514
Db		348 CCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGC	407
Qy		515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCC	574
Db		408 TGCGCAGGGACCAGTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGA	467
Qy		575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT	634
Db		468 GCAGCCTCACTGTGCGCGAGACGTTGCGATAACACAGC---GATGCTGGCCCTCTGCCGCA	524
Qy		635 TCTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC	694
Db		525 GCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCC	584
Qy		695 AGTGCCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTCTGGGGGGTGAGCGCA	754
Db		585 ACGTGGCGGACCAAATGATTGGCAGCTATAATTTGGGGGAATTTCCAGTGGCGAGCGGC	644
Qy		755 GGAGAGTCAGCATTGGGGTGCGAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC	814
Db		645 GCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGC	704
Qy		815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG	874
Db		705 CAACCACAGGACTGGACTGCATGACTGCAAATCAAATGTCTTCTCTTGCTGAGCTGG	764
Qy		875 CCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC	934
Db		765 CTCGCAGGGACCGAATTGTGATTGTCACCATCCACCAGCCTCGCTCTGAGCTCTTCCAAC	824
Qy		935 TGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGC	994
Db		825 ACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTTCTGTGGCACCCCAGAGG	884
Qy		995 ACATGGTCCAGTATTTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTG	1054
Db		885 AGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTG	944
Qy		1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA	1114
Db		945 ATTTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGT	1004
Qy		1115 GGGAGAAGGCTCAGTCACTCGCA	1137
Db		1005 ACAAGCGAGTACAGATGCTGGAA	1027

XX
AC ABK51685;
XX
DT 30-JUL-2002 (first entry)
XX
DE Mouse ABCG5 cDNA sequence.
XX
KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW ss.
XX
OS Mus sp.
XX
PN WO200227016-A2.
XX
PD 04-APR-2002.
XX
PF 25-SEP-2001; 2001WO-US029859.
XX
PR 25-SEP-2000; 2000US-0235268P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.
XX
PI Patel SB, Dean M;
XX
DR WPI; 2002-416483/44.
XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.
XX
PS Example 3; Page 45; 66pp; English.
XX
CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of the
CC invention
XX

SQ Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 U; 0 Other;

Query Match 7.2%; Score 193.4; DB 6; Length 2354;
Best Local Similarity 53.4%; Pred. No. 5e-41;
Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

```
Qy      335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
          ||  |||||  |  ||  |||  |  |  ||  |  |  |||||  |
Db      320 GCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCC 379

Qy      395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
          ||||  ||  ||||  ||||  |||||  |  ||||  ||  |  ||||  ||  |||
Db      380 AGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCT 439

Qy      455 CTGGCCGAGGTCACGGCGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
          |||  |  |  ||  |||  |  |||  |||  |  |||||
Db      440 CCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGC 499

Qy      515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCC 574
          |  |||  |  |||  |  |||||  |  |||  ||  ||||
Db      500 TGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGA 559

Qy      575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
          ||  |  |||||  |||||  |||  ||  ||  |||||  |  |  |  |
Db      560 GCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGC---GATGCTGGCCCTCTGCCGCA 616

Qy      635 TCTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
          ||||  ||  |  ||||  |  |||  ||||  ||  ||  |  |||||  |  ||  |||
Db      617 GCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCC 676

Qy      695 AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA 754
          |  ||||  |  ||||  |  |  |  ||||  ||  |  ||  |||||
Db      677 ACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGC 736

Qy      755 GGAGAGTCAGCATTTGGGGTGCGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
          |  ||||  |||  |  |  |||||  |  ||||  ||  |  ||  ||  |||
Db      737 GCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGC 796

Qy      815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG 874
          |  |||  |  ||  ||  |||  ||  ||  ||  |  |  ||  ||||  |  |||||
Db      797 CAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCTTCTCTTGGCTGAGCTGG 856

Qy      875 CCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
          |  |  |  ||||  |  |||  |  ||  ||  |||||  |||||  |||||  |
Db      857 CTCGCAGGGACCGAATTGTGATTGTCACCATCCACCAGCCTCGCTCTGAGCTCTTCCAAC 916

Qy      935 TGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCCATCTACTTAGGGGCGGCCAGC 994
          ||  ||  |  |  |  ||||  |  ||  |  ||  ||  ||  |  |  ||
Db      917 ACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTTCTGTGGCACCCAGAGG 976

Qy      995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTG 1054
          |  |||  |  |  ||||  |  ||  |||||  |||||  |  |||||  ||
Db      977 AGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCCTGAACATTCCAATCCCTTTG 1036

Qy      1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
          |  ||  ||  |||||  ||||  |  ||||  |  |||||  ||||  |  ||
Db      1037 ATTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGT 1096
```

Qy 1115 GGGAGAAGGCTCAGTCACTCGCA 1137
 || | || |
 Db 1097 ACAAGCGAGTACAGATGCTGGAA 1119

RESULT 15

ABK51687

ID ABK51687 standard; cDNA; 1069 BP.

XX

AC ABK51687;

XX

DT 07-AUG-2003 (revised)

DT 30-JUL-2002 (first entry)

XX

DE cDNA encoding hamster ABCG5 protein.

XX

KW Hamster; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ss.

XX

OS Cricetinae.

XX

FH Key Location/Qualifiers

FT CDS 30. .1049

FT /*tag= a

FT /partial

FT /product= "Hamster ABCG5 protein"

FT /note= "This sequence lacks both a start and stop codon"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR P-PSDB; AAU96987.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 47; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a

CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence encodes the hamster ABCG5 protein of the invention.
CC (Updated on 07-AUG-2003 to correct OS field.)

SQ Sequence 1069 BP; 266 A; 282 C; 273 G; 248 T; 0 U; 0 Other;

Query Match 6.5%; Score 173.2; DB 6; Length 1069;
Best Local Similarity 54.1%; Pred. No. 9.2e-36;
Matches 397; Conservative 0; Mismatches 333; Indels 4; Gaps 2;

Qy	418	TCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGCCCGAGTCCACGGCCGAAG	477
Db	1	TCAGGCTCAGGGAAAACCACGTTGCT-GGTGCCATCTCCGGGAGGCTGCGACGCACAGGG	59
Qy	478	ATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAG	537
Db	60	ACCCTGGAAGGGGAGGTGTTTGTGAACGGCCGTGAGCTGCGCAGGGACCAGTTCCAAGAC	119
Qy	538	TGTGTGGCCCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACC	597
Db	120	TGCTTCTCCTATGTCCTGCAGAGCGACGTCTTTCTGAGCAGTCTCACGGTGCGAGAGACG	179
Qy	598	TTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCACGCTGACAAA	657
Db	180	CTGCGCTACACGGCGATGCTGGCCCTCCGCAGTAGCTCTTC---GGACTTCTATGACAAG	236
Qy	658	AGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCCTGACACCCGCGTGGGC	717
Db	237	AAGGTAGAGGCAGTCATGGAAGAGCTAAGTCTGAGCCACGTGGCAGACCGAATGATTGGC	296
Qy	718	AACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTTGGGGTGCAG	777
Db	297	AACTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGCCGAGTCTCCATCGCAGCCCAA	356
Qy	778	CTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTC	837
Db	357	CTCATT CAGGACCCCAAGATCATGATGTTTGATGAGCCAACCACAGGACTGGACTGCATG	416
Qy	838	ACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTC	897
Db	417	ACTGCAAATCAAATTGTCATCCTCCTGGCAGAGCTGGCTCGCAGGGACCGCATTTGTGATC	476
Qy	898	ATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCCTCCTGATG	957

Db	477	GTCACCATCCACCAGCCTCGCTCTGAGCTCTTTCAACACTTCGACAAAATTGCCATCCTG	536
Qy	958	ACGTCTGGCACCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCC	1017
Db	537	ACTTACGGAGAGATGGTGTCTGTGGCACGCCGAGGAAATGCTCGACTTCTTCAATAGC	596
Qy	1018	ATCGGCTACCCCTGTCTCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGC	1077
Db	597	TGTGGTTACCCCTTGTCTGAACATTCGAACCCCTTTGACTTCTACATGGACTTGACATCA	656
Qy	1078	ATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCA	1137
Db	657	GTGGATACCCAGAGCAGAGAGCGAGAAATAGAAACCTACAAGAGAGTCCAGATGCTCGAA	716
Qy	1138	GCCCTGTTTCTAGA	1151
Db	717	TCTGCCTTCAGAGA	730

Search completed: February 26, 2004, 01:20:00
Job time : 680.985 secs

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:48:03 ; Search time 129.121 Seconds
(without alignments)
11471.161 Million cell updates/sec

Title: US-09-989-981A-7
Perfect score: 2669
Sequence: 1 gtgtccctgctccaggaaac.....caattaaaaatgtattgagc 2669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			%		Query				Description
	No.	Score	Match	Length	DB	ID			
	1	115.4	4.3	2418	4	US-09-245-808-2			Sequence 2, Appli
c	2	61.6	2.3	747	4	US-09-252-991A-14368			Sequence 14368, A
	3	61.6	2.3	795	4	US-09-252-991A-14337			Sequence 14337, A
c	4	61.6	2.3	1311	4	US-09-252-991A-14340			Sequence 14340, A
	5	61.6	2.3	1374	4	US-09-252-991A-14279			Sequence 14279, A
	6	61.6	2.3	4159	4	US-09-614-912-139			Sequence 139, App
	7	60.8	2.3	1977	4	US-09-614-912-143			Sequence 143, App
	8	55.8	2.1	3376	4	US-09-620-312D-918			Sequence 918, App
	9	54.6	2.0	7218	1	US-08-232-463-14			Sequence 14, Appl
	10	54	2.0	4403765	3	US-09-103-840A-2			Sequence 2, Appli
	11	54	2.0	4411529	3	US-09-103-840A-1			Sequence 1, Appli

	12	51.4	1.9	723	4	US-09-252-991A-11541	Sequence 11541, A
c	13	51.4	1.9	1155	4	US-09-252-991A-11845	Sequence 11845, A
	14	51.4	1.9	2367	4	US-09-252-991A-11600	Sequence 11600, A
c	15	51.4	1.9	28804	2	US-08-592-874-1	Sequence 1, Appli
c	16	51.4	1.9	28804	3	US-09-096-942-2	Sequence 2, Appli
c	17	51.4	1.9	28804	3	US-09-096-867-2	Sequence 2, Appli
	18	48.8	1.8	1047	4	US-08-540-650B-6	Sequence 6, Appli
	19	48.8	1.8	1053	4	US-09-016-434-1423	Sequence 1423, Ap
	20	48.8	1.8	1882	4	US-08-540-650B-11	Sequence 11, Appl
	21	48.8	1.8	3083	4	US-08-693-308-1	Sequence 1, Appli
	22	48.6	1.8	2031	4	US-09-614-912-137	Sequence 137, App
	23	48.2	1.8	900	4	US-09-252-991A-10183	Sequence 10183, A
c	24	48.2	1.8	1332	4	US-09-252-991A-9889	Sequence 9889, Ap
	25	47.2	1.8	915	4	US-09-489-039A-4785	Sequence 4785, Ap
	26	46.2	1.7	654	4	US-09-252-991A-13915	Sequence 13915, A
c	27	46.2	1.7	1509	4	US-09-252-991A-13436	Sequence 13436, A
	28	46.2	1.7	1668	4	US-09-252-991A-13705	Sequence 13705, A
c	29	45.6	1.7	435	4	US-09-252-991A-9969	Sequence 9969, Ap
	30	45.2	1.7	1458	3	US-09-105-537-9	Sequence 9, Appli
	31	45.2	1.7	1565	3	US-09-320-878-24	Sequence 24, Appl
	32	45.2	1.7	1565	4	US-09-141-908-22	Sequence 22, Appl
	33	45.2	1.7	1565	4	US-09-657-440-24	Sequence 24, Appl
	34	45.2	1.7	1698	4	US-09-252-991A-8469	Sequence 8469, Ap
c	35	45.2	1.7	1722	4	US-09-252-991A-8412	Sequence 8412, Ap
c	36	45.2	1.7	13613	3	US-09-105-537-3	Sequence 3, Appli
	37	44.8	1.7	795	4	US-09-489-039A-5947	Sequence 5947, Ap
c	38	44.6	1.7	627	4	US-09-252-991A-12021	Sequence 12021, A
c	39	44.6	1.7	705	4	US-09-252-991A-12050	Sequence 12050, A
	40	44.6	1.7	732	4	US-09-252-991A-11963	Sequence 11963, A
	41	44.6	1.7	2328	4	US-09-252-991A-11890	Sequence 11890, A
c	42	44.4	1.7	507	4	US-09-252-991A-5110	Sequence 5110, Ap
c	43	44.4	1.7	804	4	US-09-252-991A-5139	Sequence 5139, Ap
	44	44.4	1.7	1860	4	US-09-252-991A-5014	Sequence 5014, Ap
	45	44.2	1.7	876	4	US-09-489-039A-1001	Sequence 1001, Ap

ALIGNMENTS

RESULT 1

US-09-245-808-2

; Sequence 2, Application US/09245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; TITLE OF INVENTION: encodes it

; FILE REFERENCE: Ross UMb conversion

; CURRENT APPLICATION NUMBER: US/09/245,808

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-2

Query Match 4.3%; Score 115.4; DB 4; Length 2418;
Best Local Similarity 51.2%; Pred. No. 4.7e-19;
Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```
Qy      548 ACGTGC GCCAGCACAACCAGCTGCTCCCAACTTGACTGTGCGAGAGACCTTGGCCTTCA 607
      |||||  || | |  || |  ||  ||||| ||| ||||| | |||  |||
Db      606 ACGTGGTACAAGATGATGTTGTGATGGGCACTCTGACGGTGAGAGAAAACTTACAGTTCT 665

Qy      608 TTGCCCAGATGCGGCTGCCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGGGTGGAGG 667
      ||  | ||||| | | || | | |  |  || ||| ||| | |
Db      666 CAGCAGCTCTTCGGCTTGCAACAACATATGACGAATCATGAAAAAACGAACGGATTAACA 725

Qy      668 ACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTACG 727
      || ||  ||| | | ||  |  || ||| ||  || ||| |  ||
Db      726 GGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCAGACTCCAAGGTTGGAAGCTCAGTTTA 785

Qy      728 TCGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGA 787
      | || | |||| || || || | | ||  || ||| ||  || |||| |
Db      786 TCCGTGGTGTGTCTGGAGGAGAAAGAAAAAGGACTAGTATAGGAATGGAGCTTATCACTG 845

Qy      788 ACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGCCCACA 847
      | ||  ||| | | | || | || || ||| | |||  ||||| |
Db      846 ATCCTTCCATCTTGTTCTTGGATGAGCCTACAACCTGGCTTAGACTCAAGCACAGCAAATG 905

Qy      848 ACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCC 907
      | |  | | ||  ||| || | ||  ||  |  || ||||| | |
Db      906 CTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAGGGACGAACAATCATCTTCTCCATTC 965

Qy      908 ACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCA 967
      | ||||| || | | ||||| | |||||  || | | || | || | |
Db      966 ATCAGCCTCGATATTCATCTTCAAGTTGTTTGATAGCCTCACCTTATTGGCCTCAGGAA 1025

Qy      968 CCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACC 1027
      | || | | ||| | || || |  |  || ||  | |  || || |
Db      1026 GACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTGGGATACTTTGAATCAGCTGGTTATC 1085

Qy      1028 CCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGA 1072
      ||||  ||| | || ||||| |||||  |||| | |
Db      1086 ACTGTGAGGCCTATAATAACCTGCAGACTTCTTCTTGACATCA 1130
```

RESULT 2

US-09-252-991A-14368/c

; Sequence 14368, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14368
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14368

Query Match 2.3%; Score 61.6; DB 4; Length 747;
Best Local Similarity 47.6%; Pred. No. 7.6e-06;
Matches 273; Conservative 0; Mismatches 274; Indels 27; Gaps 2;

```
Qy      357 GGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAG 416
      || || || | | | | | | | | | | | | | | | | | |
Db      732 GGTGGTCAAGGGCGTCGACCTGAGGGTGGACAAGGGCGAGGTGCTGTCGATCATCGGCGG 673

Qy      417 CTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAG-----GTCACGG 470
      ||| ||| || | | | | | | | | | | | | | | | |
Db      672 CTCCGGTTCCGGCAAGTCGACCCTGCTGATGTGCATCAACGGCCTGGAGCCGATCCAGCG 613

Qy      471 CGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGT 530
      |||| | || | | | | | | | | | | | | | | |
Db      612 CGGCAGCATCCGCGTCGACGGCATCGACGTGCATGCCCCGCGGTACCGACCTCAACCGCCT 553

Qy      531 GAGGAAGTGTGTGGCCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCG 590
      | || | | | | | | | | | | | | | | | | |
Db      552 GCGGCGGAAGATCGGCATCGTCTTCCAGCAGTGAACGCCTTCCCCCACCTGACCGTGCT 493

Qy      591 AGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCG 650
      || | | | | | | | | | | | | | | | | | |
Db      492 GGAAAACGTCATGCTCGCGCCGCGCAAGGTGCTCGGCAAGAGCCGCGCCGAAGCC----- 438

Qy      651 TGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCG 710
      | | | | | | | | | | | | | | | | | |
Db      437 -----GAGGCGATGGCGCTGAAGCAACTCACCCACGTCGGTCTCGGCGA 394

Qy      711 CGTGGGCAACATGTACGTGCGGGGGTTGTCGGGGGGTGAGCGCAGGAGAGTCAGCATTGG 770
      | | | | | | | | | | | | | | | | | |
Db      393 CAAGCTCAAGGTCTTCCCCCAGCGCCTTCCGGCGGCCAGCAACAGCGCATGGCGATCGC 334

Qy      771 GGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGA 830
      | || | | | | | | | | | | | | | | | | |
Db      333 CCGGGCGCTGGCGATGTCGCCGGAATACATGCTGTTTCGACGAAGCCACCTCGGCGCTCGA 274

Qy      831 CAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCT 890
      | | | | | | | | | | | | | | | | | |
Db      273 CCCGCAGTTGGTCGGCGAGGTGGTGGACACCATGCGCATGCTCGCCGAGGAAGGCATGAC 214

Qy      891 GGTGCTCATCTCCCTCCACCAGCCTCGCTCTGAC 924
      || || | | | ||| || ||| || |
Db      213 CATGGTCCTGGTCACCCACGAGATCCGCTTCGCC 180
```

RESULT 3

US-09-252-991A-14337

; Sequence 14337, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 14337

; LENGTH: 795

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-14337

Query Match 2.3%; Score 61.6; DB 4; Length 795;
Best Local Similarity 47.6%; Pred. No. 7.8e-06;
Matches 273; Conservative 0; Mismatches 274; Indels 27; Gaps 2;

```

Qy      357 GGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAG 416
      || || || | | | | | | | | | | | | | | | | | |
Db      94 GGTGGTCAAGGGCGTCGACCTGAGGGTGGACAAGGGCGAGGTGCTGTGCATCATCGGCGG 153

Qy      417 CTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAG-----GTCACGG 470
      ||| ||| || | | | | | | | | | | | | | | | |
Db      154 CTCCGGTTCCGGCAAGTCGACCCTGCTGATGTGCATCAACGGCCTGGAGCCGATCCAGCG 213

Qy      471 CGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCAGCTCGCCTCAGCTGGT 530
      ||||| || | | | | | | | | | | | | | | | |
Db      214 CGGCAGCATCCGCGTCGACGGCATCGACGTGCATGCCCCGCGGTACCGACCTCAACCGCCT 273

Qy      531 GAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCG 590
      | || | | | | | | | | | | | | | | | | | | |
Db      274 GCGGCGGAAGATCGGCATCGTCTTCCAGCAGTGGAACGCCTTCCCCACCTGACCGTGCT 333

Qy      591 AGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCG 650
      || | | | | | | | | | | | | | | | | | | |
Db      334 GGAAAACGTCATGCTCGCGCCGCGCAAGGTGCTCGGCAAGAGCCGCGCCGAAGCC----- 388

Qy      651 TGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGGCGCTGACACCCG 710
      | | | | | | | | | | | | | | | | | | | |
Db      389 -----GAGGCGATGGCGCTGAAGCAACTACCCACGTGCGTCTCGGCGA 432

Qy      711 CGTGGGCAACATGTACGTGCGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGG 770
      | | | | | | | | | | | | | | | | | | | |
Db      433 CAAGCTCAAGGTCTTCCCCAGCGCCTTCCGGCGGCCAGCAACAGCGCATGGCGATCGC 492

Qy      771 GGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGA 830
      | || | | | | | | | | | | | | | | | | |

```

```

Db      493 CCGGGCGCTGGCGATGTCGCCGGAATACATGCTGTTTCGACGAAGCCACCTCGGGCGCTCGA 552
Qy      831 CAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCT 890
      | | | | | | | | | | | | | | | | | | | | | |
Db      553 CCCGCAGTTGGTCGGCGAGGTGGTGGACACCATGCGCATGCTCGCCGAGGAAGGCATGAC 612
Qy      891 GGTGCTCATCTCCCTCCACCAGCCTCGCTCTGAC 924
      | | | | | | | | | | | | | | | |
Db      613 CATGGTCCTGGTCACCCACGAGATCCGCTTCGCC 646

```

RESULT 4

US-09-252-991A-14340/c

; Sequence 14340, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 14340

; LENGTH: 1311

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-14340

Query Match 2.3%; Score 61.6; DB 4; Length 1311;

Best Local Similarity 47.6%; Pred. No. 9.5e-06;

Matches 273; Conservative 0; Mismatches 274; Indels 27; Gaps 2;

```

Qy      357 GGGCATCCAGAACCTAAGCTTCAAAGTGAGAAAGTGGGCAGATGCTGGCCATCATAGGGAG 416
      || || || | | | | | | | | | | | | | | | | |
Db      1207 GGTGGTCAAGGGCGTCGACCTGAGGGTGGACAAGGGCGAGGTGCTGTCGATCATCGGCGG 1148
Qy      417 CTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAG-----GTCACGG 470
      ||| |||| || | | || |||| |||| |||| | | |
Db      1147 CTCCGGTTCCGGCAAGTCGACCCTGCTGATGTGCATCAACGGCCTGGAGCCGATCCAGCG 1088
Qy      471 CGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGT 530
      |||| || | | || | | || | | | | | | |
Db      1087 CGGCAGCATCCGCGTCGACGGCATCGACGTGCATGCCC GCGGTACCGACCTCAACCGCCT 1028
Qy      531 GAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCG 590
      | || | | | | | | | | | | | | | | | |
Db      1027 GCGGCGGAAGATCGGCATCGTCTTCCAGCAGTGAACGCCTTCCCCACCTGACCGTGCT 968
Qy      591 AGAGACCTTGGCCTTCATTGCCAGATGCGGCTGCCAGAACCTTCTCCCAGGCCAGCG 650
      || | | | | | | | | | | | | | | | |
Db      967 GGAAAACGTCATGCTCGCGCCGCGCAAGGTGCTCGGCAAGAGCCGCGCCGAAGCC----- 913

```

```

Qy      651 TGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCCTGACACCCG 710
          || | ||| ||| | || || | || | || |
Db      912 -----GAGGCGATGGCGCTGAAGCAACTCACCCACGTCGGTCTCGGCGA 869

Qy      711 CGTGGGCAACATGTACGTGCGGGGGTTGTCGGGGGGTGAGCGCAGGAGAGTCAGCATTGG 770
          | | | ||| | | | | | | | | | | | | | | |
Db      868 CAAGCTCAAGGTCTTCCCCAGCGCCTTCCGGCGGCCAGCAACAGCGCATGGCGATCGC 809

Qy      771 GGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGA 830
          | ||| | | | | | | | | | | | | | | | |
Db      808 CCGGGCGCTGGCGATGTCGCCGGAATACATGCTGTTTCGACGAAGCCACCTCGGCGCTCGA 749

Qy      831 CAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCT 890
          | | | | | | | | | | | | | | | | | |
Db      748 CCCGCAGTTGGTCGGCGAGGTGGTGGACACCATGCGCATGCTCGCCGAGGAAGGCATGAC 689

Qy      891 GGTGCTCATCTCCCTCCACCAGCCTCGCTCTGAC 924
          || || | | |||| || |||| | |
Db      688 CATGGTCCTGGTCACCCACGAGATCCGCTTCGCC 655

```

RESULT 5

US-09-252-991A-14279

; Sequence 14279, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 14279

; LENGTH: 1374

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-14279

Query Match 2.3%; Score 61.6; DB 4; Length 1374;

Best Local Similarity 47.6%; Pred. No. 9.7e-06;

Matches 273; Conservative 0; Mismatches 274; Indels 27; Gaps 2;

```

Qy      357 GGGCATCCAGAACCTAAGCTTCAAAGTGAGAAAGTGGGCAGATGCTGGCCATCATAGGGAG 416
          || || || | | | | | | | | | | | | | |
Db      696 GGTGGTCAAGGGCGTCGACCTGAGGGTGGACAAGGGCGAGGTGCTGTGATCATCGGCGG 755

Qy      417 CTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAG-----GTCACGG 470
          ||| |||| || | | |||| | | | | | | | |
Db      756 CTCCGGTTCCGGCAAGTCGACCCTGCTGATGTGCATCAACGGCCTGGAGCCGATCCAGCG 815

```

Qy 471 CGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGT 530
 ||||| ||| | | ||| | | | | | |
 Db 816 CGGCAGCATCCGCGTCGACGGCATCGACGTGCATGCCCCGCGGTACCGACCTCAACCGCCT 875

Qy 531 GAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCG 590
 | || | | | | ||| ||||| | ||||| ||| ||||| |||||
 Db 876 GCGGCGGAAGATCGGCATCGTCTTCCAGCAGT'GGAACGCCTTCCCCACCTGACCGTGCT 935

Qy 591 AGAGACCTTGGCCTTCATTGCCAGATGCGGGT'GCCAGAACCTTCTCCCAGGCCAGCG 650
 || | | | | | | | | | | | | | | | | | | | |
 Db 936 GGAAAACGTCATGCTCGCGCCGCGCAAGGTGCTCGGCAAGAGCCGCGCCGAAGCC----- 990

Qy 651 TGACAAAAGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCG 710
 | | | ||| ||| | | | | | | | | | | | | | |
 Db 991 -----GAGGCGATGGCGCTGAAGCAACTCACCCACGTGGTCTCGGCGA 1034

Qy 711 CGTGGGCAACATGTACGTGCGGGGGTTGTCGGGGGGTGAGCGCAGGAGAGTCAGCATTGG 770
 | | ||| | | | | | | | | | | | | | | | | |
 Db 1035 CAAGCTCAAGGTCTTCCCCAGCGCCTTTCGGCGGCCAGCAACAGCGCATGGCGATCGC 1094

Qy 771 GGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGA 830
 | ||| | | | | | | | | | | | | | | | | | |
 Db 1095 CCGGGCGCTGGCGATGTCGCCGGAATACATGCTGTTGACGAAGCCACCTCGGCGCTCGA 1154

Qy 831 CAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCT 890
 | | | | | | | | | | | | | | | | | | | | |
 Db 1155 CCCGCAGTTGGTCGGCGAGGTGGTGGACACCATGCGCATGCTCGCCGAGGAAGGCATGAC 1214

Qy 891 GGTGCTCATCTCCCTCCACCAGCCTCGCTCTGAC 924
 || || | | |||| || |||| | |
 Db 1215 CATGGTCCTGGTCACCCACGAGATCCGCTTCGCC 1248

RESULT 6

US-09-614-912-139

; Sequence 139, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

; APPLICANT: Lee, Jian Ming

; APPLICANT: Sakai, Hajime

; APPLICANT: Weng, Zude

; APPLICANT: Caimi, Perry G

; APPLICANT: Anderson, Shawn

; TITLE OF INVENTION: Plant Metabolism Genes

; FILE REFERENCE: BB1378 US NA

; CURRENT APPLICATION NUMBER: US/09/614,912

; CURRENT FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: 60/143,401

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/143,412

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/146,650

```
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 139
; LENGTH: 4159
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-614-912-139
```

```
Query Match          2.3%; Score 61.6; DB 4; Length 4159;
Best Local Similarity 49.5%; Pred. No. 1.5e-05;
Matches 188; Conservative 0; Mismatches 189; Indels 3; Gaps 1;
```

```
Qy      697 TGC GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGGTGAGCGCAGG 756
          ||||| |||||   ||| ||| || |   | ||||| || || || || || | |
Db      411 TGC GCGGACACGATCGTCGGCGACCAGATGCAGAGGGGGATCTCCGGTGGTCAGAAGAAA 470

Qy      757 AGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCC 816
          | |||| || || | | | | | | | | | | || | || | || ||
Db      471 CGCGTCACCACCGGTGAGATGATTGTGCGGTCCAACAAAGGTTCTATTTCATGGATGAGATA 530

Qy      817 ACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTT---GTCCAGGCTG 873
          | |||| | ||||| || | | | | | | | | | | | | | |
Db      531 TCAACTGGATTGGACAGCTCCACCACATTCCAGATTGTCAAATGCCTTCAGCAAATCGTG 590

Qy      874 GCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGG 933
          | ||| | | ||||| || |||| || || | |||| | || |
Db      591 CACTTGGGCGAGGCAACCATCCTCATGTCACTCCTACAACCAGCCCCTGAGACTTTTGAG 650

Qy      934 CTGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCCATCTACTTAGGGGCGGCCCGAG 993
          || || || | |||| || | || | || | || | || || ||
Db      651 CTATTCGATGACATTATCCTACTGTCAGAAGGCCAGATTGTTTATCAGGGACCCCGCGAA 710

Qy      994 CACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCT 1053
          || | | ||| || | || | || || || || | | ||||
Db      711 TACGTCCTTGAGTTCTTTGAGTCATGCGGATTCCGCTGCCAGAGCGTAAGGGTACTGCA 770

Qy      1054 GACTTCTATGTGGACCTGAC 1073
          ||||| | ||| |||
Db      771 GACTTTCTTCAGGAGGTGAC 790
```

RESULT 7

```
US-09-614-912-143
; Sequence 143, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
```


; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Lee, Jian Ming
 ; APPLICANT: Sakai, Hajime
 ; APPLICANT: Weng, Zude
 ; APPLICANT: Caimi, Perry G
 ; APPLICANT: Anderson, Shawn
 ; TITLE OF INVENTION: Plant Metabolism Genes
 ; FILE REFERENCE: BB1378 US NA
 ; CURRENT APPLICATION NUMBER: US/09/614,912
 ; CURRENT FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,401
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,412
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/146,650
 ; PRIOR FILING DATE: 1999-07-30
 ; PRIOR APPLICATION NUMBER: 60/170,906
 ; PRIOR FILING DATE: 1999-12-15
 ; PRIOR APPLICATION NUMBER: 60/172,959
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/172,946
 ; PRIOR FILING DATE: 1999-12-21
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 143
 ; LENGTH: 1977
 ; TYPE: DNA
 ; ORGANISM: Triticum aestivum
 US-09-614-912-143

Query Match 2.3%; Score 60.8; DB 4; Length 1977;
 Best Local Similarity 47.2%; Pred. No. 1.8e-05;
 Matches 185; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

Qy	574	CCCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACC	633
Db	110	CCACATGTGACAATATATGAATCACTCGTATTTCTGCATGGCTGCGGCTTCCTGCAGAG	169
Qy	634	TTCTCCCAGGCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGG	693
Db	170	GTTGACTCAGAAAGAAGAAAGATGTTTCATCGAGGAGATCATGGATCTTGTAGAGCTCACA	229
Qy	694	CAGTGCCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGGTGAGCGC	753
Db	230	TCATTGAGGGGGGCACCTTGTGGGCTCCCTGGAGTGAATGGTCTATCAACTGAGCAACGC	289
Qy	754	AGGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAA	813
Db	290	AAGAGGCTTACAATTGCCGTGGAGCTTGTGCTAACCCGTCGATCATTTTATGGATGAG	349
Qy	814	CCCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTG	873
Db	350	CCAACATCTGGTCTTGATGCTCGTGCAGCTGCAATTGTGATGAGGACTGTTAGGAACACT	409
Qy	874	GCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGG	933
Db	410	GTTAACTGGCAGGACCGTTGTTTGCACCATCCACCAGCCAAGTATTGACATATTTGAA	469

Qy 422 GTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATCA 481

```

      |   ||| |   |   ||   ||| ||   ||| ||| |   |
Db      131 GGGCTGGCAAGTCTACATT-----CATGAACATCTTGGCAGGATACAGGGAGTCTGGAA 184
Qy      482 AGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAGTGTG 541
      |   || ||||| |   ||||| ||| |   ||   |   |||
Db      185 TGAAGGGGCAGATCCTGGTTAATGGAAGGCCACGGGAGCTGAGGACCTTCCGCAAGATGT 244
Qy      542 TGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCAACTTGACTGTGCGAGAGACCTTGG 601
      |   || |   || |   || ||||| || ||| ||| ||| ||
Db      245 CCTGCTACATCATGCAAGATGACATGCTGCTGCCGCACCTCACGGTGTTGGAAGCCATGA 304
Qy      602 CCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGGG 661
      || ||| |   || |||| |   |||| |   |   |   ||
Db      305 TGGTCTCTGCTAACCTGAAGCTGAGTGAGA-----AGCAGGAGGTGAAGAAGGAGCTGG 358
Qy      662 TGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCCTGACACCCGCGTGGGCAACA 721
      || || |   |   ||| ||| ||| ||| ||| ||| ||| ||
Db      359 TGACAGAGATCCTGACGGCACTGGGCTGATGTCGTGCTCCCACACGAGGACAGCCC--- 415
Qy      722 TGTACGTGCGGGGGTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCC 781
      |   || || || || || || || || || || || || ||
Db      416 -----TGCTCTTGCGGGGCAGAGGAAGCGTCTGGCCATCGCCCTGGAGCTGG 463
Qy      782 TGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAG 841
      |   ||||| || |   || |   || ||||| ||| ||| ||| ||
Db      464 TCAACAACCCGCCTGTCATGTTCTTTGATGAGCCCACCAGTGGTCTGGATAGCGCCTCTT 523
Qy      842 CCCACAACCTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCT 901
      |   || |||| |   || |||| |   || ||| |   || |
Db      524 GTTTCGAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGGGGGGCGGTACCATCATCTGCA 583
Qy      902 CCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCCTCCTGATGACGT 961
      || ||||| ||| |   || |||| |   |||||
Db      584 CCATCCACCAGCCCAGTGCCAAGCTCTTTGAGATGTTTGA----- 623
Qy      962 CTGGCACCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATCG 1021
      || |||| |   || |   || || || || || || || ||
Db      624 ----CAAGTGCATCTTCAAAGCGTGGTCACCAACCTGATCCCCTATCTAAAGGGACTCG 679
Qy      1022 GCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACC 1074
      ||| |   || || |||| ||| || ||||| || || || ||
Db      680 GCTTGCAATTGCCCCACCTACCACAACCCGGCTGACTTCATCATCGAGGTGGCC 732

```

RESULT 9

US-08-232-463-14

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

```

; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

```

```

Query Match          2.0%; Score 54.6; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 0.001;
Matches 15; Conservative 221; Mismatches 155; Indels 0; Gaps 0;

```

```

Qy      1425 GCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAA 1484
          | |::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1064 GATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1123

Qy      1485 CGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGA 1544
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1124 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1183

Qy      1545 GGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCC 1604
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1184 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1243

Qy      1605 GGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAACCTGAG 1664
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1244 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1303

Qy      1665 GCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTG 1724

```

```

      ::  ::::  :::::  ::  :  :::::  ::  :  :  :  :::::  ::
Db      1304 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1363
Qy      1725 CAGGATTATGGCCCTGGCCGCCGCGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTT 1784
      :  :  :  ::::  ::  :  :  :  ::::  :::::  :::::  :  :  :::::  ::
Db      1364 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1423
Qy      1785 CAGCAATGCCCTCTACAACCTCTTCTACCTC 1815
      :  :  :  ::::  |||  |  |||||  |||
Db      1424 YYYYYYYYGTACCAAATCTTCTATCTC 1454

```

RESULT 10

US-09-103-840A-2

```

; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g

```

US-09-103-840A-2

```

Query Match          2.0%; Score 54; DB 3; Length 4403765;
Best Local Similarity 49.0%; Pred. No. 0.018;
Matches 187; Conservative 0; Mismatches 180; Indels 15; Gaps 1;

```

```

Qy      528 GGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGT 587
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      1965623 GCTGCGCAGCAGGATCGGCATGGTGCCACAGGACGACGTGGTGCACGGTCAGCTGACCGT
1965682
Qy      588 GCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCCA 647
      | | | | | | | | | | | | | | | | | | | | | |
Db      1965683 GAAACACGCGCTGATGTATGCCGCCGAACCTACGGCTGCCGCCGGACACCACCAAAGATGA
1965742
Qy      648 GCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACAC 707
      || | | | | | | | | | | | | | | | | | | | |
Db      1965743 CCGCACCCAGGTAGTTGCCCGGGTGCTCGAAGAACTCGAGATGTCCAAGCACATCGACAC
1965802

```

Qy 708 CCGCGTGGGCAACATGTACGTGCGGGGGTTGTCTGGGGGGTGAGCGCAGGAGAGTCAGCAT 767
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1965803 CAGGGTCGACAA-----GCTGTCTGGGTGGTCAACGCAAGCGGGCGTCGGT
 1965847

Qy 768 TGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCT 827
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1965848 GGCGCTTGAGCTGTTGACCGGGCCGTCCTGCTGATCCTCGACGAGCCGACATCCGGCCT
 1965907

Qy 828 CGACAGCTTCACAGCCCAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCG 887
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1965908 AGATCCTGCGCTGGACCGGCAGGTCATGACCATGCTGCGGCAGTTGGCCGACGCCGGTCG
 1965967

Qy 888 GCTGGTGCTCATCTCCCTCCAC 909
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1965968 GGTGGTGCTCGTGGTTACCCAC 1965989

RESULT 11

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match 2.0%; Score 54; DB 3; Length 4411529;
 Best Local Similarity 49.0%; Pred. No. 0.018;
 Matches 187; Conservative 0; Mismatches 180; Indels 15; Gaps 1;

Qy 528 GGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGT 587
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1974794 GCTGCGCAGCAGGATCGGCATGGTGCCACAGGACGACGTGGTGCACGGTCAGCTGACCGT
 1974853

Qy 588 GCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCCA 647
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1974854 GAAACACGCGCTGATGTATGCCGCCGAACCTACGGCTGCCGCCGGACACCACCAAAGATGA
 1974913

Qy	648	GCGTGACAAAAGGGTGGAGGACGTGATCGGGAGCTGCGGCTTAGGCAGTGCGCTGACAC	707
Db	1974914	CCGCACCCAGGTAGTTGCCCGGGTGCTCGAAGAACTCGAGATGTCCAAGCACATCGACAC	
	1974973		
Qy	708	CCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCAT	767
Db	1974974	CAGGGTCGACAA-----GCTGTGCGGTGGTCAACGCAAGCGGGCGTCGGT	
	1975018		
Qy	768	TGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCT	827
Db	1975019	GGCGCTTGAGCTGTTGACCGGGCCGTCACTGCTGATCCTCGACGAGCCGACATCCGGCCT	
	1975078		
Qy	828	CGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCG	887
Db	1975079	AGATCCTGCGCTGGACCGGCAGGTCATGACCATGCTGCGGCAGTTGGCCGACGCCGGTCTG	
	1975138		
Qy	888	GCTGGTGCTCATCTCCCTCCAC	909
Db	1975139	GGTGGTGCTCGTGGTTACCCAC	1975160

```

      ||| ||      |      | | | | |      |      | | | | |      |
Db      205 CGTGGGGTCTCGGCCACGATGCGAGGAGCCTGGAGGCGCGCCGCCAGCTCGGCTACCT 264

Qy      588 GCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCCA 647
      || ||| | | | ||| | | | | |      |      | | | | |      |
Db      265 GCCGGAAGACGTGACCTTCTACCCGAGCTCAGCGGCGCGGAAACCTGCGCCACTTCGC 324

Qy      648 GCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACAC 707
      || ||| | | | | | | | | | | | | | | | | | | | |
Db      325 CCGCCTCAAGGGCGTGGCGCCGCGCCGAAGCCGCGCGCTGCTGGAACAGGTGCGCCTCGG 384

Qy      708 CCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGTGAGCGCAGGAGAGTCAGCAT 767
      || | | | | | | | | | | | | | | | | | | | | |
Db      385 CCATGCAGCCAGGCGGCGCTGAAAACCTACTCGAAGGGCATGCGCCAGCGCCTCGGCCT 444

Qy      768 TGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCT 827
      |      | ||| || | | | | | | | | | | | | | | | |
Db      445 GGCCCAGGCGCTGCTCGGCGAACC GCGCCTGCTGCTGCTCGACGAACCGACGGTGGGCCT 504

Qy      828 CGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTG---GCCAAAGGCAA 884
      ||| | | | | | | | | | | | | | | | | | | |
Db      505 CGACCCGCTGGCCACCGTCGAGCTCTACCAATTGCTCGACCGCCTGCGCGGCCAGGGCAC 564

Qy      885 CCGGCTGGTGTCTATCTCCCTCCACCAGCCTCGCTCTGACA 925
      | | | | | | | | | | | | | | | | | | | | |
Db      565 CGGGATCGTCCTTTGCTCCCATGTGCTGCCCCGGCGTCGAGA 605

```

RESULT 13

US-09-252-991A-11845/c

; Sequence 11845, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 11845

; LENGTH: 1155

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11845

Query Match 1.9%; Score 51.4; DB 4; Length 1155;

Best Local Similarity 46.0%; Pred. No. 0.0031;

Matches 212; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

```

Qy      468 CGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCT 527
      || ||| ||| || | | | | | | | | | | | | |

```



```

Db      859 CGCCGGCAAGACCACCACCATCAAGCTGGTCTCTCGGCCTGCTGGCCCCAGCGAAGGCCG 800
Qy      528 GGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGT 587
      ||| || | | | | | | | | | | | | | | |
Db      799 CGTGCGGGTCCTCGGCCACGATGCGAGGAGCCTGGAGGCGCGCCGCGCAGCTCGGCTACCT 740
Qy      588 GCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCCA 647
      || |||| | | |||| | ||| | | |||| | ||| |
Db      739 GCCGGAGAACGTGACCTTCTACCCGCAGCTCAGCGGCGCGGAAACCTGCGCCACTTCGC 680
Qy      648 GCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACAC 707
      || ||| | |||| | | || | || | ||| ||| || |
Db      679 CCGCCTCAAGGGCGTGGCGCCGGCCGAAGCCGCGCGCCTGCTGGAACAGGTGCGCCTCGG 620
Qy      708 CCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCAT 767
      || | | || | || | || | || |||| | | ||| |
Db      619 CCATGCAGCCAGGCGGCGCCTGAAAACCTACTCGAAGGGCATGCGCCAGCGCCTCGGCCT 560
Qy      768 TGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCT 827
      | | | ||| || | || | | || | ||||| || | |||
Db      559 GGCCCAGGCGCTGCTCGGCGAACCGCGCCTGCTGCTGCTCGACGAACCGACGGTGGGCCT 500
Qy      828 CGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTG---GCCAAAGGCAA 884
      |||| | | || | || | | | ||| | | ||| |||
Db      499 CGACCCGCTGGCCACCGTCGAGCTCTACCAATTGCTCGACCGCCTGCGCGGCCAGGGCAC 440
Qy      885 CCGGCTGGTGTCTATCTCCCTCCACCAGCCTCGCTCTGACA 925
      | || | || || |||| | ||| || || |
Db      439 CGGGATCGTCCTTTGCTCCCATGTGCTGCCCGGCGTCGAGA 399

```

RESULT 14

US-09-252-991A-11600

; Sequence 11600, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 11600

; LENGTH: 2367

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11600

Query Match 1.9%; Score 51.4; DB 4; Length 2367;

Best Local Similarity 46.0%; Pred. No. 0.0042;

Matches 212; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

Qy 468 CGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCT 527
 || ||||| || | | | | || || || | |
 Db 1569 CGCCGGCAAGACCACCACCATCAAGCTGGTCCTCGGCCTGCTGGCCCCCAGCGAAGGCCG 1628

Qy 528 GGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGT 587
 ||| || | | | | | | | | | | | | |
 Db 1629 CGTGCGGGTCCTCGGCCACGATGCGAGGAGCCTGGAGGCGCGCCGCCAGCTCGGCTACCT 1688

Qy 588 GCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCCA 647
 || |||| | | |||| | ||| | | | |||| | || |
 Db 1689 GCCGGAGAACGTGACCTTCTACCCGAGCTCAGCGGCGCGGAAACCCTGCGCCACTTCGC 1748

Qy 648 GCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACAC 707
 || || | |||| | | || | | | || | || | || |
 Db 1749 CCGCCTCAAGGGCGTGGCGCCGGCCGAAGCCGCGCGCTGCTGGAACAGGTGCGCCTCGG 1808

Qy 708 CCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCAT 767
 || | | || | | || | || | || | || | || |
 Db 1809 CCATGCAGCCAGGCGGCGCCTGAAAACCTACTCGAAGGGCATGCGCCAGCGCCTCGGCCT 1868

Qy 768 TGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCT 827
 | | | || || | | || | | | | ||||| || || ||
 Db 1869 GGCCCAGGCGCTGCTCGGCGAACCGCGCCTGCTGCTGCTCGACGAACCGACGGTGGGCCT 1928

Qy 828 CGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTG---GCCAAAGGCAA 884
 |||| | | || | || | | | | | || | | ||||
 Db 1929 CGACCCGCTGGCCACCGTCGAGCTCTACCAATTGCTCGACCGCCTGCGCGGCCAGGGCAC 1988

Qy 885 CCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACA 925
 | || | || || |||| | ||| || || |
 Db 1989 CGGGATCGTCCTTTGCTCCCATGTGCTGCCCCGGCGTCGAGA 2029

RESULT 15

US-08-592-874-1/c

; Sequence 1, Application US/08592874

; Patent No. 5854034

; GENERAL INFORMATION:

; APPLICANT: POLLOCK, THOMAS J.

; APPLICANT: YAMAZAKI, MOTOHIDE

; APPLICANT: THORNE, LINDA

; APPLICANT: MIKOLAJCZAK, MARCIA

; APPLICANT: ARMENTROUT, RICHARD W.

; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING

; TITLE OF INVENTION: POLYSACCHARIDE PRODUCTION

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JULES E. GOLDBERG

; STREET: 261 MADISON AVENUE

; CITY: NEW YORK

; STATE: NY

; COUNTRY: USA

; ZIP: 10016-2391

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

```

;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,874
;      FILING DATE:
;      CLASSIFICATION:  435
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/377,440
;      FILING DATE:  24-JAN-1995
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  GOLDBERG, JULES E.
;      REGISTRATION NUMBER:  24,408
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-986-4090
;      TELEFAX:  212-818-9479
;      INFORMATION FOR SEQ ID NO:  1:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  28804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;      MOLECULE TYPE:  DNA (genomic)
;      FRAGMENT TYPE:  N-terminal
US-08-592-874-1

```

```

Query Match          1.9%;  Score 51.4;  DB 2;  Length 28804;
Best Local Similarity  47.9%;  Pred. No. 0.011;
Matches 148;  Conservative  0;  Mismatches 161;  Indels  0;  Gaps  0;

```

```

Qy      606 CATTGCCAGATGCGGCTGCCAGAACCTTCTCCCAGGCCAGCGTGACAAAAGGGTGGGA 665
        ||||| | | | | | | | | | | | | | | | | | | | |
Db      19351 CATTGCGCTGTCCAACCCGGCGATGCCGTTTCGAGCATGTCGTGGCGGCGGCGACGCTGGC 19292

Qy      666 GGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTA 725
        || | | | | | | | | | | | | | | | | | | |
Db      19291 GGGTGCATGACTTCATCCTGCGTCAGCCGCGCGGCTATGACACCGAGATCGTCGAGCG 19232

Qy      726 CGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCCTGTG 785
        || | | | | | | | | | | | | | | | | | | |
Db      19231 CGGCGTCAACCTGTGCGGCGGCCAGCGCCAGCGGCTCGCTATCGCCCGCGCGCTGGTCGG 19172

Qy      786 GAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGCCCA 845
        || | | | | | | | | | | | | | | | | | | |
Db      19171 CAATCCGCGCATCCTGGTGTTCGACGAGGCGACCTCCGCGCTGGATGCCGAGAGCGAGGA 19112

Qy      846 CAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCT 905
        | | | | | | | | | | | | | | | | | | | |
Db      19111 GCTGATCCAGAACAATCTGCGCGCCATCTCGGCGGGCCGCGACGCTGGTGATCATCGCCA 19052

Qy      906 CCACCAGCC 914
        || | | | |
Db      19051 CCGCCTGTC 19043

```

```

Search completed: February 26, 2004, 09:46:10
Job time : 144.121 secs

```

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 06:40:42 ; Search time 613.323 Seconds
(without alignments)
15698.623 Million cell updates/sec

Title: US-09-989-981A-7
Perfect score: 2669
Sequence: 1 gtgtccctgctccaggaaac.....caattaaaaatgtattgagc 2669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2353733 seqs, 1803733377 residues

Total number of hits satisfying chosen parameters: 4707466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		%				
Result		Query				
No.	Score	Match	Length	DB	ID	Description

1	2669	100.0	2669	10	US-09-989-981A-7	Sequence 7, Appli
2	1680.6	63.0	3239	15	US-10-415-378-29	Sequence 29, Appl
3	1430	53.6	2019	10	US-09-989-981A-3	Sequence 3, Appli
4	203.6	7.6	2340	9	US-09-837-992-4	Sequence 4, Appli
5	203.6	7.6	2340	10	US-09-989-981A-5	Sequence 5, Appli
6	193.4	7.2	1959	10	US-09-989-981A-1	Sequence 1, Appli
7	193.4	7.2	2258	9	US-09-837-992-2	Sequence 2, Appli
8	176.8	6.6	2585	12	US-10-425-114-32175	Sequence 32175, A
9	122	4.6	2025	9	US-09-866-866A-13	Sequence 13, Appl
10	115.6	4.3	972	12	US-10-424-599-129897	Sequence 129897,
11	115.4	4.3	2027	15	US-10-405-806-1	Sequence 1, Appli
12	115.4	4.3	2053	15	US-10-405-806-12	Sequence 12, Appl
13	115.4	4.3	2247	9	US-09-866-866A-26	Sequence 26, Appl
14	115.4	4.3	2418	10	US-09-961-086-2	Sequence 2, Appli
15	115.4	4.3	2574	9	US-09-981-353-34	Sequence 34, Appl
16	115.4	4.3	2718	14	US-10-120-687-60	Sequence 60, Appl
17	115.4	4.3	2719	9	US-09-866-866A-9	Sequence 9, Appli
18	115.4	4.3	2883	14	US-10-101-510-639	Sequence 639, App
19	115	4.3	2608	12	US-10-424-599-33340	Sequence 33340, A
20	114	4.3	925	12	US-10-424-599-64406	Sequence 64406, A
21	109	4.1	2512	15	US-10-104-047-825	Sequence 825, App
22	106.8	4.0	3201	13	US-10-072-621-5	Sequence 5, Appli
23	106.8	4.0	3463	12	US-10-425-114-13286	Sequence 13286, A
24	106.8	4.0	3528	12	US-10-424-599-99236	Sequence 99236, A
25	105.2	3.9	2930	9	US-09-954-531-591	Sequence 591, App
26	105.2	3.9	2930	14	US-10-171-581-276	Sequence 276, App
27	105.2	3.9	2930	16	US-10-429-160-9	Sequence 9, Appli
c 28	103.2	3.9	567	15	US-10-260-238-5734	Sequence 5734, Ap
29	103.2	3.9	2826	12	US-10-424-599-95809	Sequence 95809, A
c 30	101.4	3.8	447	9	US-09-960-352-11649	Sequence 11649, A
31	100.8	3.8	2028	15	US-10-369-493-27486	Sequence 27486, A
32	98.2	3.7	2133	15	US-10-210-130-13	Sequence 13, Appl
33	97.4	3.6	2182	10	US-09-949-029-23	Sequence 23, Appl
34	94.6	3.5	2162	12	US-10-425-114-25098	Sequence 25098, A
35	94.6	3.5	2248	12	US-10-425-114-31870	Sequence 31870, A
36	93.8	3.5	1684	12	US-10-425-114-4362	Sequence 4362, Ap
c 37	93.6	3.5	427	9	US-09-960-352-12839	Sequence 12839, A
38	93.6	3.5	2400	13	US-10-108-605-244	Sequence 244, App
39	93.6	3.5	2638	15	US-10-369-493-27088	Sequence 27088, A
40	92.2	3.5	2788	9	US-09-745-763-196	Sequence 196, App
41	92	3.4	10330	13	US-10-001-189-68	Sequence 68, Appl
42	91.8	3.4	452	12	US-10-424-599-107241	Sequence 107241,
c 43	90.6	3.4	6043	10	US-09-989-981A-9	Sequence 9, Appli
c 44	90.2	3.4	377	9	US-09-960-352-4326	Sequence 4326, Ap
c 45	89.4	3.3	414	9	US-09-960-352-5514	Sequence 5514, Ap

ALIGNMENTS

RESULT 1

US-09-989-981A-7

; Sequence 7, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

```
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2669
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (100)..(2121)
; OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981A-7
```

```
Query Match          100.0%; Score 2669; DB 10; Length 2669;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 GTGTCCCTGCTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTGTCCCTGCTCCAGGAAACAGAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGTCT 60

Qy     61 AAGAGAGCTGCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGGAAGGCGGCAGAG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGAGCTGCAGCCCAGGGTCACAGACCTGTGGGCCCCATGGCCGGGAAGGCGGCAGAG 120

Qy    121 GAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACCTCGGGCCTCCAGGATAGATTG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACCTCGGGCCTCCAGGATAGATTG 180

Qy    181 TTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTACAGTGGCCAGCCCAACACCCTG 240

Qy    241 GAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCTCAGGTCCCTTGGTTTGAGCAG 300

Qy    301 CTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGCCAGAATTCTTGTGAGCTGGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGCCAGAATTCTTGTGAGCTGGGC 360

Qy    361 ATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCA 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATGCTGGCCATCATAGGGAGCTCA 420

Qy    421 GGTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATC 480
```

Db	421		GGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGCCGAGGTCACGGCGGCAAGATC	480
Qy	481		AAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAGTGT	540
Db	481		AAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCGCCTCAGCTGGTGAGGAAGTGT	540
Qy	541		GTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTG	600
Db	541		GTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTG	600
Qy	601		GCCTTCATTGCCAGATGCGGCTGCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGG	660
Db	601		GCCTTCATTGCCAGATGCGGCTGCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGG	660
Qy	661		GTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAAC	720
Db	661		GTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAAC	720
Qy	721		ATGTACGTGCGGGGGTTGTCGGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTC	780
Db	721		ATGTACGTGCGGGGGTTGTCGGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTC	780
Qy	781		CTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACA	840
Db	781		CTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACA	840
Qy	841		GCCCAACCTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATC	900
Db	841		GCCCAACCTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATC	900
Qy	901		TCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACG	960
Db	901		TCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACG	960
Qy	961		TCTGGCACCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATC	1020
Db	961		TCTGGCACCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATC	1020
Qy	1021		GGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATT	1080
Db	1021		GGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATT	1080
Qy	1081		GACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCC	1140
Db	1081		GACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCC	1140
Qy	1141		CTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGAT	1200
Db	1141		CTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGAT	1200
Qy	1201		CTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCG	1260
Db	1201		CTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCG	1260
Qy	1261		AGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATT	1320

Db	1261	AGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACGACGCTGATCCGTCTGCAGATT	1320
Qy	1321	TCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATG	1380
Db	1321	TCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATG	1380
Qy	1381	TCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGAT	1440
Db	1381	TCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGAT	1440
Qy	1441	ACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTC	1500
Db	1441	ACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTC	1500
Qy	1501	ATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTAC	1560
Db	1501	ATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTAC	1560
Qy	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTAC	1620
Db	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTAC	1620
Qy	1621	ATCATCATCTACGGGATGCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCC	1680
Db	1621	ATCATCATCTACGGGATGCCACCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCC	1680
Qy	1681	TTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTG	1740
Db	1681	TTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTG	1740
Qy	1741	GCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTAC	1800
Db	1741	GCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTAC	1800
Qy	1801	AACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCC	1860
Db	1801	AACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCC	1860
Qy	1861	GCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTTCAG	1920
Db	1861	GCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTTCAG	1920
Qy	1921	TTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGAT	1980
Db	1921	TTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGAT	1980
Qy	1981	AAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTC	2040
Db	1981	AAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTC	2040
Qy	2041	ATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAG	2100
Db	2041	ATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAG	2100
Qy	2101	AAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGC	2160
Db	2101	AAACCAAGTCAAGACTGGTGATTACGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGC	2160

Qy 2161 AGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCTTCTGTTGGGACAGTGAGGACAATGA 2220
 |||
 Db 2161 AGACCCTTCAACTGCACTCCCTCCTCAGGAGCCCCTTCTGTTGGGACAGTGAGGACAATGA 2220

Qy 2221 CCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAG 2280
 |||
 Db 2221 CCCTACAGATGCTCAGCTACATCCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAG 2280

Qy 2281 GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTCTGCTCACTGGCAGGAGACTGCGAT 2340
 |||
 Db 2281 GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTCTGCTCACTGGCAGGAGACTGCGAT 2340

Qy 2341 GACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGA 2400
 |||
 Db 2341 GACTGGGAGAAAACCTGCACTCGGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGA 2400

Qy 2401 TATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAG 2460
 |||
 Db 2401 TATGCATTTATATAGGCAACTCGATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAG 2460

Qy 2461 CTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTG 2520
 |||
 Db 2461 CTAGACTGTGCAGGAATTGTTGGAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTG 2520

Qy 2521 GCTTCATCTTCCAGGGGGCCCCCACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACC 2580
 |||
 Db 2521 GCTTCATCTTCCAGGGGGCCCCCACTCCGTGGTGAGCCACCATCAATACAGAAAGTGACC 2580

Qy 2581 TAAGATGTACCAGCAAGATGCCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAAAGC 2640
 |||
 Db 2581 TAAGATGTACCAGCAAGATGCCATCCCTTCTTTTGTGTGGGGTCATGGGCTCCAAAAGC 2640

Qy 2641 CAACGTGAACAATTAAAAATGTATTGAGC 2669
 |||
 Db 2641 CAACGTGAACAATTAAAAATGTATTGAGC 2669

RESULT 2

US-10-415-378-29

; Sequence 29, Application US/10415378

; Publication No. US20040014945A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom
 ; APPLICANT: YUE, Henry; NGUYEN, Dannel B.;
 ; APPLICANT: HAFALIA, April J.A.; ELLIOTT, Vicki S.;
 ; APPLICANT: LU, Yan; CHAWLA, Narinder K.;
 ; APPLICANT: YAO, Monique G.; BAUGHN, Mariah R.;
 ; APPLICANT: GANDHI, Ameena R.; DING, Li;
 ; APPLICANT: SANJANWALA, Madhusudan M.; RAMKUMAR, Jayalaxmi;
 ; APPLICANT: ARVIZU, Chandra S.; GIETZEN, Kimberly J.;
 ; APPLICANT: LAL, Preeti G.; AZIMZAI, Yalda;
 ; APPLICANT: KHAN, Farrah A.; THANGAVELU, Kavitha;
 ; APPLICANT: THORNTON, Michael B.; LU, Dyung Aina M.;
 ; APPLICANT: TRIBOULEY, Catherine M.; WARREN, Bridget A.;
 ; APPLICANT: ISON, H. Craig; DAS, Debopriya;
 ; APPLICANT: RAUMANN, Brigitte E.; POLICKY, Jennifer L.;

Qy	1343	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	1402
Db	372	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	431
Qy	1403	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA	1462
Db	432	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA	491
Qy	1463	TCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGA	1522
Db	492	TCGGTGCTCTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGA	551
Qy	1523	GGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG	1582
Db	552	GGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG	611
Qy	1583	CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCATCTACGGGATGCCCA	1642
Db	612	CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCATCTACGGGATGCCCA	671
Qy	1643	CCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGG	1702
Db	672	CCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGG	731
Qy	1703	TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA	1762
Db	732	TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA	791
Qy	1763	CCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGG	1822
Db	792	CCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGG	851
Qy	1823	GCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCT	1882
Db	852	GCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCT	911
Qy	1883	TCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTAGTTTCAGCAGAAGAACTTATAAAA	1942
Db	912	TCCTGCGGTGGTGTGTTTGAAGGGCTGATGAAGATTAGTTTCAGCAGAAGAACTTATAAAA	971
Qy	1943	TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC	2002
Db	972	TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC	1031
Qy	2003	TGGAATCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA	2062
Db	1032	TGGAATCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA	1091
Qy	2063	TGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGAT	2122
Db	1092	TGGTCCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGAT	1151
Qy	2123	TCACGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGCAGACCCCTTCAACTGCACTCCCT	2182
Db	1152	TCACGCCAGACGTCTGCCCGCTGGTGGGGGACCTGAGCAGACCCCTTCAACTGCACTCCCT	1211

Qy	2183	CCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACAT	2242
Db	1212	CCTCAGGAGCCCCCTTCCTGGGGACAGTGAGGACAATGACCCTACAGATGCTCAGCTACAT	1271
Qy	2243	CCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAG	2302
Db	1272	CCGGCCCAGGGTGCTGCAGTGGCACAGACCAGCCACAGGATGGCAGTAGAATAAAGACAG	1331
Qy	2303	TCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTC	2362
Db	1332	TCGAAAGGGATTTCTGCTCACTGGCAGGAGACTGCGATGACTGGGAGAAAACCTGCACTC	1391
Qy	2363	GGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATATAGGCAACTC	2422
Db	1392	GGTGGCACCTACAACGTTGCTAATTTATTTCTTTTGATATGCATTTATATAGGCAACTC	1451
Qy	2423	GATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTG	2482
Db	1452	GATATAGGATGGGAGCAAACCTAGGAATGAATTGGGTAGCTAGACTGTGCAGGAATTGTTG	1511
Qy	2483	GAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCAGGGGCCCCA	2542
Db	1512	GAACCTGGAGGGAACAATAACAGTAGCTAGCAGATTTGGCTTCATCTTCAGGGGCCCCA	1571
Qy	2543	CCTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATGCC	2602
Db	1572	CCTCCGTGGTGAGCCACCATCAATACAGAAAGTGACCTAAGATGTACCAGCAAGATGCC	1631
Qy	2603	ATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAATTAAAAATGT	2662
Db	1632	ATCCCTTCTTTTTGTGTGGGGTCATGGGCTCCAAAAGCCAACGTGAACAATTAAAAATGT	1691
Qy	2663	ATTGAGC	2669
Db	1692	ATTGAGC	1698

RESULT 3

US-09-989-981A-3

; Sequence 3, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

; FILE REFERENCE: 018781-007320US

; CURRENT APPLICATION NUMBER: US/09/989,981A

; CURRENT FILING DATE: 2002-07-23

; PRIOR APPLICATION NUMBER: US 60/252,235

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/253,645

; PRIOR FILING DATE: 2000-11-28

; NUMBER OF SEQ ID NOS: 13

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2019)
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-3
```

```
Query Match          53.6%; Score 1430; DB 10; Length 2019;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 1659; Conservative 0; Mismatches 360; Indels 3; Gaps 1;
```

```
Qy      100 ATGGCCGGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACC 159
        ||||| | ||| | ||| |||| |||| ||| ||| | | ||||| |
Db       1  ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60

Qy      160 TCGGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTAC 219
        ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db       61 TCGGGCCTCCAGGACAGCTTGTTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120

Qy      220 AGTGGCCAGCCCAACACCCTGGAGGTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCT 279
        ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180

Qy      280 CAGGTCCCTTGTTTGTAGCAGCTGGCTCAGTTCAGATGCCCTGGACATCTCCCAGCTGC 339
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db      181 CAGGTGCCTTGTTTGTAGCAGCTGGCTCAGTTCAGATACCCTGGAGGTCTCATAGCAGC 240

Qy      340 CAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATG 399
        || | || ||||| ||||| || ||||| ||||| ||||| |||||
Db      241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300

Qy      400 CTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGC 459
        ||||| ||||| ||||| || ||||| || || || ||||| |||
Db      301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360

Qy      460 CGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCG 519
        |||| |||| ||||| || |||| || || |||| ||||| |||| ||
Db      361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420

Qy      520 CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAAC 579
        ||||| ||||| || || || |||| |||| |||| ||||| |||||
Db      421 CCTCAGCTGGTGAGGAAGTGCCTTGCATGTGCGGCAGCATGACCAACTGCTGCCCAAC 480

Qy      580 TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCC 639
        |||| || ||||| |||| ||||| ||||| ||||| ||||| |||||
Db      481 CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC 540

Qy      640 CAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGC 699
        ||||| ||||| ||||| |||| |||| ||||| ||||| |||||
Db      541 CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC 600

Qy      700 GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGGTGAGCGCAGGAGA 759
        || ||||| | ||||| || || || || || ||||| ||||| || ||
```

Db 601 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA 660
 Qy 760 GTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC 819
 || |||||
 Db 661 GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 720
 Qy 820 TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAA 879
 |||||
 Db 721 TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCTGGCCAAG 780
 Qy 880 GGCAACCGGCTGGTGTCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTT 939
 |||||
 Db 781 GGCAACAGGCTGGTGTCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT 840
 Qy 940 GATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACATG 999
 || |||||
 Db 841 GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG 900
 Qy 1000 GTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTC 1059
 || |||||
 Db 901 GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC 960
 Qy 1060 TATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAG 1119
 || |||||
 Db 961 TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG 1020
 Qy 1120 AAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTGACTTAGATGACTTTCTA 1179
 |||||
 Db 1021 AAGGCACAGTCTCTTGCAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG 1080
 Qy 1180 TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCA 1239
 |||||
 Db 1081 TGGAAAGCTGAGGCAAAGGAACCTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA 1140
 Qy 1240 CTAGACACCAACTGCCTCCCGAGTCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG 1299
 | |||||
 Db 1141 CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTC 1197
 Qy 1300 ACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCCACCCTCCTCATCCAT 1359
 || |||||
 Db 1198 ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTTCAT 1257
 Qy 1360 GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGC 1419
 ||| |||||
 Db 1258 GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCC 1317
 Qy 1420 ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTCATGATCGGTGCTCTCATCCCT 1479
 | |||||
 Db 1318 AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCCT 1377
 Qy 1480 TTCAACGTCACTTCTGGATGTCTCCAAATGTACTCAGAGAGGGCAATGCTTTACTAT 1539
 |||||
 Db 1378 TTCAATGTCTCCTGGATGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT 1437
 Qy 1540 GAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG 1599
 || |||||
 Db 1438 GAGCTGGAAGACGGGCTGTACACTGCTGGTCCCTTATTTCTTTGCCAAGATCCTAGGAGAA 1497


```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human sitosterolemia gene (SSG)
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: protein
US-09-837-992-4
```

```
Query Match          7.6%; Score 203.6; DB 9; Length 2340;
Best Local Similarity 54.4%; Pred. No. 2.6e-51;
Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;
```

```
Qy      335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
          ||| ||| | | | | | | | | | | | | | | | | | | | |
Db      285 GCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGC 344

Qy      395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
          ||| | | | | | | | | | | | | | | | | | | | |
Db      345 AGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGT 404

Qy      455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
          | | | | | | | | | | | | | | | | | | | | | |
Db      405 CCGGGAGGCTGGGGCGCGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGCGC 464

Qy      515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACAGCTGCTCC 574
          | | | | | | | | | | | | | | | | | | | | |
Db      465 TGCGCCGGGAGCAGTTCAGGACTGCTTCTCTACGTCCTGCAGAGCGACACCCTGCTGA 524

Qy      575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
          ||| | | | | | | | | | | | | | | | | | | |
Db      525 GCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCCGGGCA 584

Qy      635 TCTCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
          || | | | | | | | | | | | | | | | | | | | |
Db      585 ATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCC 641

Qy      695 AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA 754
          | | | | | | | | | | | | | | | | | | | | |
Db      642 ATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGC 701

Qy      755 GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
          | | | | | | | | | | | | | | | | | | | | |
Db      702 GCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGC 761

Qy      815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG 874
          | | | | | | | | | | | | | | | | | | | | |
Db      762 CAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGAAGTGG 821

Qy      875 CCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
          | | | | | | | | | | | | | | | | | | | | |
Db      822 CTCGCAGGAACCGAATTGTGGTTCTCACCATTACACAGCCCCGTCTGAGCTTTTTTCAGC 881
```


Qy 935 TGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGC 994
 | | | | | | | | | | | | | | | | | | | | | |
 Db 882 TCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGG 941
 Qy 995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTG 1054
 | | | | | | | | | | | | | | | | | | | | | |
 Db 942 AAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCCTGAACATTCAAACCCCTTTTG 1001
 Qy 1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1002 ACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCT 1061
 Qy 1115 GGGAGAAGGCTCAG 1128
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1062 CCAAGAGAGTCCAG 1075

RESULT 5

US-09-989-981A-5

; Sequence 5, Application US/09989981A
 ; Publication No. US20030049730A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hobbs, Helen H.
 ; APPLICANT: Shan, Bei
 ; APPLICANT: Barnes, Robert
 ; APPLICANT: Tian, Hui
 ; APPLICANT: Tularik Inc.
 ; APPLICANT: Board of Regents, The University of Texas System
 ; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
 ; FILE REFERENCE: 018781-007320US
 ; CURRENT APPLICATION NUMBER: US/09/989,981A
 ; CURRENT FILING DATE: 2002-07-23
 ; PRIOR APPLICATION NUMBER: US 60/252,235
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/253,645
 ; PRIOR FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 2340
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (107)..(2062)
 ; OTHER INFORMATION: human ABCG5 (hABCG5)
 US-09-989-981A-5

Query Match 7.6%; Score 203.6; DB 10; Length 2340;
 Best Local Similarity 54.4%; Pred. No. 2.6e-51;
 Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

Qy 335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
 | | | | | | | | | | | | | | | | | | | | | |
 Db 285 GCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGC 344
 Qy 395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454

```

      |||| | | |||| |||| ||||| || | |||| | | | |||| | | | ||
Db      345 AGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGT 404
Qy      455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
      | | | | | || | || | || | || | | | | | | | | | |
Db      405 CCGGGAGGCTGGGGCGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGCGC 464
Qy      515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCC 574
      | | | | | | | | | | | | | | | | | | | | | |
Db      465 TGCGCCGGGAGCAGTTCCAGGACTGCTTCTCTACGTCCTGCAGAGCGACACCCTGCTGA 524
Qy      575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
      | | | | | | | | | | | | | | | | | | | | | |
Db      525 GCAGCCTCACCGTGCGCGAGACGCTGCACTACACGCGCTGCTGGCCATCCGCCGCGGCA 584
Qy      635 TCTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
      || | | | | | | | | | | | | | | | | | | | |
Db      585 ATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCC 641
Qy      695 AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA 754
      | | | | | | | | | | | | | | | | | | | | | |
Db      642 ATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGC 701
Qy      755 GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
      | | | | | | | | | | | | | | | | | | | | | |
Db      702 GCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGC 761
Qy      815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG 874
      | | | | | | | | | | | | | | | | | | | | | |
Db      762 CAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCTGCTCCTCTGGTGAAGTGG 821
Qy      875 CCAAAGGCAACCGGCTGGTGTCTATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
      | | | | | | | | | | | | | | | | | | | | | |
Db      822 CTCGCAGGAACCGAATTGTGGTTCTCACCATTCAACAGCCCCGTTCTGAGCTTTTTTCAGC 881
Qy      935 TGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCCATCTACTTAGGGGCGGCCAGC 994
      | | | | | | | | | | | | | | | | | | | | | |
Db      882 TCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGG 941
Qy      995 ACATGGTCCAGTATTTTCACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTG 1054
      | | | | | | | | | | | | | | | | | | | | | |
Db      942 AAATGCTTGATTCTTCAATGACTGCGGTTACCCCTGTCTGAACATTCAAACCCTTTTG 1001
Qy      1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
      | | | | | | | | | | | | | | | | | | | | | |
Db      1002 ACTTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCT 1061
Qy      1115 GGGAGAAGGCTCAG 1128
      || | | |
Db      1062 CCAAGAGAGTCCAG 1075

```

RESULT 6

US-09-989-981A-1

; Sequence 1, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

```
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1959)
; OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-1
```

```
Query Match          7.2%; Score 193.4; DB 10; Length 1959;
Best Local Similarity 53.4%; Pred. No. 3.3e-48;
Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;
```

```
Qy      335 GCTGCCAGAATTC TTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
      ||  |||| |      ||  ||| | | | ||| | | |||| |
Db      182 GCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGACATCGAGAGTGGCC 241

Qy      395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
      |||| || |||| |||| ||||| | |||| || | |||| || | |||
Db      242 AGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCT 301

Qy      455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
      ||| | | | || || || | | ||| || | | |||||
Db      302 CCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGC 361

Qy      515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCC 574
      | | | | | | | | || | || ||| | || | || | ||
Db      362 TGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGA 421

Qy      575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
      || | | ||||| |||| || | || || |||| | | | | |
Db      422 GCAGCCTCACTGTGCGGAGACGTTGCGATACACAGC---GATGCTGGCCCTCTGCCGCA 478

Qy      635 TCTCCCAGGCCAGCGTGACAAAAGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
      |||| || | |||| | || ||| || || | ||||| | || || |
Db      479 GCTCCGCGGACTTCTACAACAAGAAGGTAGAGGACGTCATGACAGAGCTGAGCCTGAGCC 538

Qy      695 AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA 754
      | ||| ||| | |||| | | | |||| | || | || ||||
Db      539 ACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGC 598
```

Qy 755 GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
 | | | | | | | | | | | | | | | | | | | | | |
 Db 599 GCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGC 658
 Qy 815 CCACCTCTGGGCTCGACAGCTTCACAGCCCAACCTGGTGAAGACCTTGTCAGGCTGG 874
 | | | | | | | | | | | | | | | | | | | | | |
 Db 659 CAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCCTTCTCTGGCTGAGCTGG 718
 Qy 875 CCAAAGGCAACCGGCTGGTGCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
 | | | | | | | | | | | | | | | | | | | | | |
 Db 719 CTCGCAGGGACCGAATTGTGATTGTCAACATCCACCAGCCTCGCTCTGAGCTCTTCCAAC 778
 Qy 935 TGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGC 994
 | | | | | | | | | | | | | | | | | | | | | |
 Db 779 ACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCAGAGG 838
 Qy 995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTG 1054
 | | | | | | | | | | | | | | | | | | | | | |
 Db 839 AGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTG 898
 Qy 1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
 | | | | | | | | | | | | | | | | | | | | | |
 Db 899 ATTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGT 958
 Qy 1115 GGGAGAAGGCTCAGTCACTCGCA 1137
 | | | | | | | | | | | | | | | | | | | | | |
 Db 959 ACAAGCGAGTACAGATGCTGGAA 981

RESULT 7

US-09-837-992-2

; Sequence 2, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2258

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)

; NAME/KEY: CDS

; LOCATION: (47)..(2005)
; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: protein
US-09-837-992-2

Query Match 7.2%; Score 193.4; DB 9; Length 2258;
Best Local Similarity 53.4%; Pred. No. 3.5e-48;
Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

```
Qy      335 GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC 394
      || ||||| | || | || | | | || | | ||||| |
Db      228 GCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCC 287

Qy      395 AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA 454
      |||| | | |||| |||| ||||| | |||| | | |||| | | |||
Db      288 AGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCT 347

Qy      455 CTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA 514
      | || | | || || || | | || | || | | |||||
Db      348 CCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGC 407

Qy      515 GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCC 574
      | | | | | || | | || || | || || | || | | | |
Db      408 TGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGA 467

Qy      575 CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCT 634
      || | | ||||| ||||| || | || || ||||| | | | | |
Db      468 GCAGCCTCACTGTGCGGAGACGTTGCGATACACAGC---GATGCTGGCCCTCTGCCGCA 524

Qy      635 TCTCCCAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC 694
      |||| | | | |||| | || |||| | || | | ||||| | || | |
Db      525 GCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCC 584

Qy      695 AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCA 754
      | || |||| | |||| | | | |||| | || | |||||
Db      585 ACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGC 644

Qy      755 GGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAAC 814
      | |||| || | | |||| | |||| | || || || || || |
Db      645 GCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGC 704

Qy      815 CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGG 874
      | || | | | || || || || || | | || | |||| | |||||
Db      705 CAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCTTCTCTTGGCTGAGCTGG 764

Qy      875 CCAAAGGCAACCGGCTGGTGTCTATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934
      | | | |||| | || | || || ||||| ||||| ||||| |
Db      765 CTCGCAGGGACCGAATTGTGATTGTCAACCATCCACCAGCCTCGCTCTGAGCTCTTCCAAC 824

Qy      935 TGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCCATCTACTTAGGGGCGGCCCAGC 994
      || || | | | |||| | || | || || || | | ||
Db      825 ACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTTCTGTGGCACCCCAGAGG 884

Qy      995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTG 1054
      | || | | | |||| | || ||||| ||||| ||||| ||
Db      885 AGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTG 944
```

```

Qy      1055 ACTTCTATGTGGACCTGACCAGCATTTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      945 ATTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGT 1004

Qy      1115 GGGAGAAGGCTCAGTCACTCGCA 1137
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1005 ACAAGCGAGTACAGATGCTGGAA 1027

```

RESULT 8

US-10-425-114-32175

; Sequence 32175, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 32175

; LENGTH: 2585

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: UC-ZMFLB73274A02_FLI

US-10-425-114-32175

Query Match 6.6%; Score 176.8; DB 12; Length 2585;

Best Local Similarity 55.3%; Pred. No. 4.9e-43;

Matches 368; Conservative 0; Mismatches 292; Indels 6; Gaps 1;

```

Qy      391 GGGCAGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTG 450
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      605 GGGTCGCTGACCGCGCTCATGGGGCCCTCGGGGTCCGGCAAGTCCACCCTGCTCGACGCC 664

Qy      451 ATCACTGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAG 510
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      665 CTCGCCGGCCGCCCTCGCCGCCAACGCCTTCCTCTCCGGCAACGTGCTCCTCAACGG---- 720

Qy      511 CCCAGCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACAGCTG 570
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      721 --CCGCAAGGCCAAGCTCTCCTTCGGCGCCGCGGCGTACGTGACGCAGGACGACAACCTG 778

Qy      571 CTCCCCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGA 630
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      779 ATCGGGACGCTGACGGTGCAGGACGATCGGCTACTCGGCGCTGCTGCGGCTGCCGGAC 838

Qy      631 ACCTTCTCCCAGGCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTT 690
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	839	AAGATGCCGCGGGAGGACAAGCGCGCGCTGGTGGAGGGCACCATCGTCGAGATGGGGCTG	898
Qy	691	AGGCAGTGCCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTCGGGGGGGTGAG	750
Db	899	CAGGACTGCGCCGACACCGTCATCGGCAACTGGCACCTCCGCGGGGTGAGCGGCGGCGAG	958
Qy	751	CGCAGGAGAGTCAAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGAC	810
Db	959	AAGCGCCGCGTCAAGCATCGCGCTCGAGCTACTCATGCGCCCGCGCCTCCTCTTCTCTCGAC	1018
Qy	811	GAACCCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGG	870
Db	1019	GAGCCCACCGAGCGGCCTCGACAGCTCGTCTGCGTTCTTCGTGACGCAGACGCTGCGGGGC	1078
Qy	871	CTGGCCAAAGGCAACCGGCTGGTGTCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTC	930
Db	1079	CTGGCGAGGGACGGCAGGACGGTGATTGCTTCCATCCACCAGCCAGCAGCGAGGTGTTC	1138
Qy	931	AGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCCATCTACTTAGGGGGCGGCC	990
Db	1139	GAGCTCTTCGACATGCTCTTCTGCTATCCGGGGGCAAGACCGTCTACTTCGGACAAGCA	1198
Qy	991	CAGCACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCTGCTACAGCAATCCT	1050
Db	1199	TCGCAAGCATGCGAGTTCTTTGCTCAAGCCGGTTCCCTTGCCCGGCTCTGCGGAATCCG	1258
Qy	1051	GCTGAC	1056
Db	1259	TCCGAC	1264

US-09-866-866A-13

; Patent No. US20020102244A1

; GENERAL INFORMATION:

; APPLICANT: Sorrentino, Brian

; APPLICANT: Schuetz, John

; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells

; FILE REFERENCE: 1340-1-021CIP2

; CURRENT APPLICATION NUMBER: US/09/866,866A

; CURRENT FILING DATE: 2001-08-30

; PRIOR APPLICATION NUMBER: 09/584,586

; PRIOR FILING DATE: 2000-05-31

; PRIOR APPLICATION NUMBER: PCT/US99/11825

; PRIOR FILING DATE: 1999-05-27

; PRIOR APPLICATION NUMBER: 60/086,988

; PRIOR FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 27

```
; SOFTWARE: PatentIn version 3.0
```

; SEQ ID NO 13

; LENGTH: 2025

; TYPE: DNA

; ORGANISM: Mus musculus

US-09-866-866A-13

Query Match

Best Local Similarity 51.7%; Pred. No. 3.1e-26;
Matches 278; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

```
Qy      538 TGTGTGGCCACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACC 597
      ||| | | ||| || | || | | ||| ||| ||| |||
Db      374 TGTTTCAGGTTATGTGGTTCAAGATGACGTTGTGATGGGCACCCTGACAGTGAGAGAAAAC 433

Qy      598 TTGGCCTTCATTGCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCGTGACAAA 657
      || ||| || | || ||| ||| ||| | | ||
Db      434 TTACAGTTCTCAGCAGCTCTTCGACTTCCAACAACATATGAAGAATCATGAAAAAATGAA 493

Qy      658 AGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGC 717
      || | | | || || ||| | || | | ||| || | |||
Db      494 CGGATTAACACAATCATTAAGAGTTAGGTCTGGAAAAAGTAGCAGATTCTAAGGTCGGA 553

Qy      718 AACATGTACGTGCGGGGGTGTGCGGGGGTGAGCGCAGGAGAGT'GAGCATTGGGGTGCAG 777
      | || | ||| || | || ||| || | || | ||| ||| ||| |||
Db      554 ACTCAGTTTATCCGTGGCATCTCTGGAGGAGAAAGAAAAAGGACAAGCATAGGGATGGAG 613

Qy      778 CTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGGCTCGACAGCTTC 837
      || | | ||| |||| | || ||| |||| |||| | ||| |
Db      614 CTGATCACTGACCCTTCCATCCTCTTCTGGATGAGCCACGACTGGTTTGGACTCAAGC 673

Qy      838 ACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTC 897
      |||| | | | | | || || || || || || || || || ||
Db      674 ACAGCGAATGCTGTCCTTTTGCTCCTGAAAAGGATGTCTAAACAGGGTCGAACAATCATC 733

Qy      898 ATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCCTCCTGATG 957
      |||| | || |||| || | |||| | | |||| || || | ||
Db      734 TTCTCCATTATCAGCCTCGGTATTCATCTTTAAGTTGTTTGACAGCCTCACCTTACTG 793

Qy      958 ACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTTCACAGCC 1017
      | || || | | | | | || | || || | | | |||| || || |
Db      794 GCTTCCGGGAAACTCGTGTTCCATGGGCCAGCACAGAAGGCCCTTGGAGTACTTTGCATCA 853

Qy      1018 ATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCA 1075
      || |||| |||| | |||| || |||| || || | || | || |
Db      854 GCAGGTTACCACTGTGAGCCCTACAACAACCCTGCGGATTTTTCTTGTATGTATCA 911
```

RESULT 10

US-10-424-599-129897

; Sequence 129897, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 129897

; LENGTH: 972
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_88305C.1
US-10-424-599-129897

Query Match 4.3%; Score 115.6; DB 12; Length 972;
Best Local Similarity 53.7%; Pred. No. 2e-24;
Matches 288; Conservative 0; Mismatches 239; Indels 9; Gaps 2;

```
Qy      549 CGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTGGCCTTCAT 608
      ||| | ||| || || | ||| || | || | ||| ||| || ||| |
Db      401 CGTTCCCCAAGACGACGTTCACTACCCTCACCTCACAGTGTTAGAGACTTTAACCTACGC 460

Qy      609 TGCCCAGATGCGGCTGCCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGGGTGGAGGA 668
      || | ||| ||| || | || | || | || | || | |||
Db      461 AGCGTTATTGAGACTTCCGAAGAGTTTGAGCAGAGAAGAGAAGGAGCACGCGGAGAT 520

Qy      669 CGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGG-----GCAACAT 722
      ||||| ||||| ||| | | || | || | || | ||| |||
Db      521 GGTGATTGCGGAGCTAGGGCTAACACGGTGTCGTAACAGCCCCGTTGGAGGGTGCATGGC 580

Qy      723 GTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCCT 782
      | ||| | |||| | || | || | | |||| | || | || | |
Db      581 TCTGTTCCGTGGCATTTCGGGTGGGGAACGGAACGGGTCACTATCGGGCAGGAGATGTT 640

Qy      783 GTGGAACCCAGGAATCCTTATTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGC 842
      | |||| | | | || | || | ||||| || | ||| | ||| ||
Db      641 GGTCAACCCGAGTTTGTGTGTTGTGATGAGCCACCTCGGGCTTGGACTCCACCACGGC 700

Qy      843 CCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTC 902
      ||| | ||| | | | |||| || | || || ||| ||
Db      701 CCAACTTATTGTGTGCGGTGCTCCGCGGGCTCGCCTTGGCGGGTCGAACCGTCGTCACCAC 760

Qy      903 CCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTC 962
      | ||||| ||| | | || | ||| ||||| || | || ||| |
Db      761 CATCCACCAGCCCTCCAGCCGGTTGTATAGGATGTTTGATAAGGTGGTCGTGTTGTCAGA 820

Qy      963 TGGCACCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATCGG 1022
      ||| ||| || || || || || | | | ||| || | || ||||
Db      821 TGGGTACCCAATTTATAGCGGGCAGGCGGGTCGGGTCATGGACTATCTCGGATCCGTCGG 880

Qy      1023 CTA---CCCCTGTCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGACCA 1075
      || ||| | || || || || || || || || |||| | |
Db      881 ATATGTCCAGCTTTCAACTTCATGAACCCAGCAGATTTCCTTCTTGACCTTGCTA 936
```

RESULT 11

US-10-405-806-1

; Sequence 1, Application US/10405806
; Publication No. US20030232362A1
; GENERAL INFORMATION:
; APPLICANT: KOMATANI, HIDEYA
; APPLICANT: HARA, YOSHIKAZU
; APPLICANT: KOTANI, HIDEHITO
; APPLICANT: NAKAGAWA, RINAKO

```

; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF
; FILE REFERENCE: 234985US0CONT
; CURRENT APPLICATION NUMBER: US/10/405,806
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: PCT/JP01/08112
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP2000-303441
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2027
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45)..(2009)
US-10-405-806-1

```

```

Query Match          4.3%; Score 115.4; DB 15; Length 2027;
Best Local Similarity 51.2%; Pred. No. 3.3e-24;
Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```

```

Qy      548 ACGTGCGCCAGCACAACCAGCTGCTCCCCAAGTTGACTGTGCGAGAGACCTTGGCCTTCA 607
        |||||  || | |  || |  ||  ||||| ||| ||||| | |||  |||
Db      412 ACGTGGTACAAGATGATGTTGTGATGGGCACTCTGACGGTGAGAGAAAACCTTACAGTTCT 471

Qy      608 TTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCCGAGCGTGACAAAAGGGTGGAGG 667
        ||  | ||||| | | ||| | | |  |  || ||| || | |
Db      472 CAGCAGCTCTTCGGCTTGCAACAACCTATGACGAATCATGAAAAAACGAACGGATTAACA 531

Qy      668 ACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTACG 727
        || ||  ||| | | ||  |  || ||| ||  || ||| |  ||
Db      532 GGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCAGACTCCAAGGTTGGAACCTCAGTTTA 591

Qy      728 TGCGGGGGTTGTCTGGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGA 787
        | ||| ||| || || || | | ||  || ||| ||  || |||| |
Db      592 TCCGTGGTGTGTCTGGAGGAGAAAGAAAAAGGACTAGTATAGGAATGGAGCTTATCACTG 651

Qy      788 ACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGCCCACA 847
        | ||  ||| | | | || || || || ||| | |||  ||||| |
Db      652 ATCCTTCCATCTTGTCTTGGATGAGCCTACAACCTGGCTTAGACTCAAGCACAGCAAATG 711

Qy      848 ACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCC 907
        | |  | | ||  ||| || ||  ||  ||  | || ||||| ||
Db      712 CTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAGGACGAACAATCATCTTCTCCATTC 771

Qy      908 ACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCA 967
        | ||||| || | ||||| || |||||  || | | || | || |
Db      772 ATCAGCCTCGATATTCATCTTCAAGTTGTTTGATAGCCTCACCTTATTGGCCTCAGGAA 831

Qy      968 CCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACC 1027
        | ||| | ||| || ||| |  |  || ||  | | |||| |
Db      832 GACTTATGTTCCACGGGCTGCTCAGGAGGCTTGGGATACTTTGAATCAGCTGGTTATC 891

Qy      1028 CCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGA 1072
        ||||  ||| | || |||| |||||  |||| | |

```


Qy 848 ACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCC 907
 | | | | | | | | | | | | | | | | | | | | | |
 Db 699 CTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAGGGACGAACAATCATCTTCTCCATTC 758
 Qy 908 ACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCA 967
 | | | | | | | | | | | | | | | | | | | | | |
 Db 759 ATCAGCCTCGATATTCCATCTTCAAGTTGTTTGATAGCCTCACCTTATTGGCCTCAGGAA 818
 Qy 968 CCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACC 1027
 | | | | | | | | | | | | | | | | | | | | | |
 Db 819 GACTTATGTTCCACGGGCTGCTCAGGAGGCCTTGGGATACTTTGAATCAGCTGGTTATC 878
 Qy 1028 CCTGTCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGA 1072
 | | | | | | | | | | | | | | | | | | | | | |
 Db 879 ACTGTGAGGCCTATAATAACCCTGCAGACTTCTTCTTGACATCA 923

RESULT 13

US-09-866-866A-26

; Sequence 26, Application US/09866866A
 ; Patent No. US20020102244A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sorrentino, Brian
 ; APPLICANT: Schuetz, John
 ; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
 ; FILE REFERENCE: 1340-1-021CIP2
 ; CURRENT APPLICATION NUMBER: US/09/866,866A
 ; CURRENT FILING DATE: 2001-08-30
 ; PRIOR APPLICATION NUMBER: 09/584,586
 ; PRIOR FILING DATE: 2000-05-31
 ; PRIOR APPLICATION NUMBER: PCT/US99/11825
 ; PRIOR FILING DATE: 1999-05-27
 ; PRIOR APPLICATION NUMBER: 60/086,988
 ; PRIOR FILING DATE: 1998-05-28
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 26
 ; LENGTH: 2247
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-866-866A-26

Query Match 4.3%; Score 115.4; DB 9; Length 2247;
 Best Local Similarity 51.2%; Pred. No. 3.5e-24;
 Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

Qy 548 ACGTGCGCCAGCACAAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTGGCCTTCA 607
 | | | | | | | | | | | | | | | | | | | | | |
 Db 561 ACGTGGTACAAGATGATGTTGTGATGGGCACTCTGACGGTGAGAGAAAACCTTACAGTTCT 620
 Qy 608 TTGCCCAGATGCGGCTGCCCAGAACCTTCTCCAGGCCAGCGTGACAAAAGGGTGGAGG 667
 | | | | | | | | | | | | | | | | | | | | | |
 Db 621 CAGCAGCTCTTCGGCTTGCAACAATATGACGAATCATGAAAAAACGAACGGATTAAACA 680
 Qy 668 ACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTACG 727
 | | | | | | | | | | | | | | | | | | | | | |
 Db 681 GGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCAGACTCCAAGGTTGGAACCTCAGTTTA 740

Qy 728 TGC GGGGGT TGT C GGGGGT GAG CGC AGG AGA GTC AGC ATT GGG GTG CAG CT CCT GTG GA 787
 | | | | | | | | | | | | | | | | | | | |
 Db 741 TCC GTGGT GTG TCT G GAG GAG AAA AGA AAA AGG ACT AGT ATA GGA ATG GAG CT TAT CACT G 800

Qy 788 ACC CAG GAAT CCT TAT TCT CGA CAA CCC ACCT CTG GG CTG AC AG CT TCA CAG CCC ACA 847
 | | | | | | | | | | | | | | | | | | | |
 Db 801 ATC CT TCC AT CT TGT TCT TGG ATG AG CCT ACA ACT GG CT TAG ACT CA AG CAC AG CAA ATG 860

Qy 848 ACCT GGT GA AG AC CT TGT CC AGG CT G GCC AA AGG CA ACC GGT GGT GCT CAT CT C CCT CC 907
 | | | | | | | | | | | | | | | | | | | |
 Db 861 CTGT CCT TTT GCT CCT G AAA AGG ATG TCT AAG CAG GG AC GA ACA AT CAT CT TCT CCA TTC 920

Qy 908 ACC AG CCT CG CT CTG AC AT CT TCA GG CT G TTT GAT CT GGT CCT CCT GAT GAC GT CT G GCA 967
 | | | | | | | | | | | | | | | | | | | |
 Db 921 ATC AG CCT CG AT AT TCC AT CT TCA AG TT G TTT GAT AG CCT CAC CT TAT TGG CCT CAG GAA 980

Qy 968 CCCCC AT CT ACT TAG GGG CGG CCC AGC AC AT GGT CC AGT AT TTT CAC AG CC AT C GG CT ACC 1027
 | | | | | | | | | | | | | | | | | | | |
 Db 981 GACT TAT GT TCC AG GGC CTG CT CAG GAG GC CT TGG GATA CT TTT GA AT CAG CT GGT TAT C 1040

Qy 1028 CCT GT CCT CG CT AC AG CA AT CCT GCT GACT TCT AT GT G GAC CT GA 1072
 | | | | | | | | | | | | | | | | | | | |
 Db 1041 ACT GT GAG GC CT ATA ATA ACC CTG CAG ACT TCT TCT TGG AC AT CA 1085

RESULT 14

US-09-961-086-2

; Sequence 2, Application US/09961086
 ; Publication No. US20030036645A1
 ; GENERAL INFORMATION:
 ; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
 ; APPLICANT: ROSS, Douglas D.
 ; APPLICANT: DOYLE, L. Austin
 ; APPLICANT: ABRUZZO, Lynne
 ; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
 ; TITLE OF INVENTION: WHICH ENCODES IT
 ; FILE REFERENCE: EP19376-019
 ; CURRENT APPLICATION NUMBER: US/09/961,086
 ; CURRENT FILING DATE: 2001-09-21
 ; PRIOR APPLICATION NUMBER: US 60/073,763
 ; PRIOR FILING DATE: 1998-02-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/02577
 ; PRIOR FILING DATE: 1999-02-05
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2418
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-961-086-2

Query Match 4.3%; Score 115.4; DB 10; Length 2418;
 Best Local Similarity 51.2%; Pred. No. 3.7e-24;
 Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

Qy 548 ACGT GCG CCAG CACA ACCAG CTG CT CCCCCA ACTTG ACTGT GCG AGAG ACCTTGG CCTTCA 607

```

      |||||  ||  |  |  ||  |  ||  ||||  |||  |||  |||
Db      606 ACGTGGTACAAGATGATGTTGTGATGGGCACTCTGACGGTGAGAGAAAACCTACAGTTCT 665
Qy      608 TTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCGTGACAAAAGGGTGGAGG 667
      ||  |||||  |  |||  |  |  |  |  ||||  |||  |
Db      666 CAGCAGCTCTTCGGCTTGCAACAACATGACGAATCATGAAAAAACGAACGGATTAACA 725
Qy      668 ACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTACG 727
      ||  ||  ||||  |  |||  |  |  |  |  ||||  ||  ||
Db      726 GGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCAGACTCCAAGGTTGGAACCTCAGTTTA 785
Qy      728 TCGGGGGGTTGTGCGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGACGCTCCTGTGGA 787
      |  ||  ||  ||||  ||  |||  |  |  |  |  ||  ||  ||  ||||  |
Db      786 TCCGTGGTGTGTCTGGAGGAGAAAGAAAAGGACTAGTATAGGAATGGAGCTTATCACTG 845
Qy      788 ACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGCCCACA 847
      |  ||  ||||  |  |  |  ||  ||  ||  ||||  |  |||  |||||  |
Db      846 ATCCTTCCATCTTGTTCTTGGATGAGCCTACAACCTGGCTTAGACTCAAGCACAGCAAATG 905
Qy      848 ACCTGGTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCC 907
      |  |  |  |  ||  ||||  ||  ||  ||  ||  ||  ||  ||||  ||
Db      906 CTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAGGGACGAACAATCATCTTCTCCATTC 965
Qy      908 ACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCA 967
      |  |||||  ||  |  |||||  |  |||||  |  ||||  ||  ||  ||  ||
Db      966 ATCAGCCTCGATATTCCATCTTCAAGTTGTTTGATAGCCTCACCTTATTGGCCTCAGGAA 1025
Qy      968 CCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACC 1027
      |  ||  ||  ||||  ||  |||  |  |  |  |  ||  ||  ||  ||||  |
Db      1026 GACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTGGGATACTTTGAATCAGCTGGTTATC 1085
Qy      1028 CCTGTCTCTGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGA 1072
      ||||  |||  |  ||  ||||  |||||  ||||  ||  |
Db      1086 ACTGTGAGGCCTATAATAACCCTGCAGACTTCTTCTTGGACATCA 1130

```

RESULT 15

US-09-981-353-34

; Sequence 34, Application US/09981353

; Patent No. US20020160382A1

; GENERAL INFORMATION:

; APPLICANT: Lasek, Amy W.

; APPLICANT: Jones, David A.

; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER

; FILE REFERENCE: PA-0038 US

; CURRENT APPLICATION NUMBER: US/09/981,353

; CURRENT FILING DATE: 2001-10-11

; NUMBER OF SEQ ID NOS: 194

; SOFTWARE: PERL Program

; SEQ ID NO 34

; LENGTH: 2574

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CB1

US-09-981-353-34

Query Match 4.3%; Score 115.4; DB 9; Length 2574;
 Best Local Similarity 51.2%; Pred. No. 3.8e-24;
 Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

```

Qy      548 ACGTGCGCCAGCACAACCAGCTGCTCCCCAACTTGACTGTGCGAGAGACCTTGGCCTTCA 607
        |||||  ||  |  |  ||  |  ||  ||||| ||| |||||  |||  |||
Db      776 ACGTGGTACAAGATGATGTTGTGATGGGCACTCTGACGGTGAGAGAAAACCTTACAGTTCT 835

Qy      608 TTGCCCAGATGCGGCTGCCCAGAACCTTCTCCCAGGCCAGCGTGACAAAAGGGTGGAGG 667
        ||      | |||||  | | |||  |  |  |  |  |  || || || | |
Db      836 CAGCAGCTCTTCGGCTTGCAACAACCTATGACGAATCATGAAAAAAACGAACGGATTAACA 895

Qy      668 ACGTGATCGCGGAGCTGCGGCTTAGGCAGTGCGCTGACACCCGCGTGGGCAACATGTACG 727
        || ||      ||| |  | ||  |  || ||| ||  || ||| |  ||
Db      896 GGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCAGACTCCAAGGTTGGAACCTCAGTTTA 955

Qy      728 TGC GGGGGTTGTGCGGGGGGTGAGCGCAGGAGAGTCAGCATTGGGGTGCAGCTCCTGTGGA 787
        | || ||  |||| | | || |  | | ||  || || | | || |||| |
Db      956 TCCGTGGTGTGTCTGGAGGAGAAAGAAAAAGGACTAGTATAGGAATGGAGCTTATCACTG 1015

Qy      788 ACCCAGGAATCCTTATTCTCGACGAACCCACCTCTGGGCTCGACAGCTTCACAGCCCACA 847
        | ||  ||| |  | | || | | || | |||  | |||  |||||  |
Db     1016 ATCCTTCCATCTTGTTCTTGGATGAGCCTACAACCTGGCTTAGACTCAAGCACAGCAAATG 1075

Qy      848 ACCTGGTGAAGACCTTGTCAGGCTGGCCAAAGGCAACCGGCTGGTGCTCATCTCCCTCC 907
        |  |  |  | ||  ||| || | ||  ||      |  |  || |||| | |
Db     1076 CTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAGGGACGAACAATCATCTTCTCCATTC 1135

Qy      908 ACCAGCCTCGCTCTGACATCTTCAGGCTGTTTGATCTGGTCCTCCTGATGACGTCTGGCA 967
        | ||||| || |  ||||| || | |||||  ||  |  |  || | || | |
Db     1136 ATCAGCCTCGATATTCCATCTTCAAGTTGTTTGATAGCCTCACCTTATTGGCCTCAGGAA 1195

Qy      968 CCCCCATCTACTTAGGGGCGGCCAGCACATGGTCCAGTATTTACAGCCATCGGCTACC 1027
        |  || |  ||| | || || |  |  || ||  |  |  || || |
Db     1196 GACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTGGGATACTTTGAATCAGCTGGTTATC 1255

Qy     1028 CCTGTCCTCGCTACAGCAATCCTGCTGACTTCTATGTGGACCTGA 1072
        ||||      ||| |  || |||| | |||||  |||||  | |
Db     1256 ACTGTGAGGCCTATAATAACCCTGCAGACTTCTTCTTGGACATCA 1300

```

Search completed: February 27, 2004, 07:11:42
 Job time : 616.323 secs